

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 01-16-03
Searcher: Beverly C 4994
Terminal time: 25
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG Suite
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
_____ ☒ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 16, 2003, 08:51:46 ; Search time 57 Seconds
(without alignments)
1028.601 Million cell updates/sec

Title: US-09-830-111a-2

Perfect score: 2194

Sequence: 1 MASPALRIRSSRSIASLR.....ESPAKALEQLTKVLTISR 440

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1_ /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194	100.0	440	22	AA1980.DAT
2	1243	56.7	451	23	AA1981.DAT
3	1064	48.5	502	23	AA1982.DAT
4	1028.5	46.9	537	23	AA1997
5	999	45.5	425	23	AA14998
6	996	45.4	403	23	AA14999
7	902.5	41.1	378	23	AA14998
8	863	39.3	377	18	AA130762
9	742	33.8	430	22	AA198466
10	658	30.0	321	21	AA19412

11	658	30.0	321	23	AB181705
12	658	30.0	321	23	AB181705
13	628.5	28.6	272	23	AB181763
14	561	25.6	297	21	AB19411
15	561	25.6	297	23	AB181704
16	561	25.6	297	23	AB181704
17	451.5	20.6	323	22	AA134728
18	451.5	20.6	323	22	AA134728
19	451.5	20.6	323	22	AA134728
20	445.5	20.3	329	22	AA135549
21	445.5	20.3	329	22	AA135549
22	445.5	20.3	329	22	AA135549
23	445.5	20.3	329	22	AA135549
24	439.5	20.0	335	21	AA121095
25	410	18.7	323	17	AA192060
26	409	18.6	320	19	AA174722
27	407.5	18.6	448	22	AB165155
28	404.5	18.5	322	22	AA136438
29	404.5	18.4	325	23	AA121871
30	402.5	18.3	325	20	AA101634
31	398.5	18.2	333	22	AA175107
32	397.5	18.1	333	23	AA121866
33	396.5	18.1	410	19	AA153921
34	395.5	18.0	337	23	AA121867
35	395	18.0	315	19	AA153922
36	395	18.0	315	23	AA121868
37	395	18.0	348	19	AA153920
38	393.5	17.9	320	23	AA121870
39	378.5	17.3	348	22	AA151415
40	371.5	16.9	322	22	AA137010
41	368	16.8	319	23	AB154681
42	368	16.8	538	21	AA181868
43	367.5	16.8	312	22	AA133819
44	362.5	16.5	321	23	AB149219
45	357	16.3	326	23	AB149754

ALIGNMENTS

RESULT 1
AAB74623
ID AAB74623 standard; Protein; 440 AA.
XX
AC AAB74623;
XX
DT 23-MAY-2001 (first entry)
XX
DE Saitoella complicata decaprenyl diphosphate synthase protein SEQ.2.
XX
KW Saitoella complicata; decaprenyl diphosphate synthase; coenzyme Q10;
XX fungus.
XX
OS Saitoella complicata.
XX
PN WO200114567-A1.
XX
PD 01-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-JP05659.
XX
PR 24-AUG-1999; 99JP-0237561.
XX
(KANF) KANEKA CORP.
XX
PI Matsuda H, Kawamukai M, Yajima K, Ikenaka Y, Hasegawa J;
XX Takahashi S;
XX
DR WPI; 2001-202937/20.
XX
PT N-PSDB; AAF74893.
XX
PT DNA encoding a protein having decaprenyl diphosphate synthase activity
and microorganism for producing coenzyme Q10 -

Arabidopsis sp. AT
Arabidopsis cocoph
Human ovarian anti
A straight-chain c
Arabidopsis sp. AT
Arabidopsis cocoph
E. coli cellular p
Salmonella typhi c
Haemophilus influe
Isoprenoid related
Paracoccus denitri
Paracoccus denitri
Agrobacterium KMK7
Heptaprenyl diphos
Bacillus stearothe
Drosophila melanog
Pseudomonas aerugi
Isoprenoid related
A decaprenyl diphos
Rhodobacter sphaer
Isoprenoid related
Decaprenyl diphosp
Isoprenoid related
Propionibacterium
Staphylococcus aur
Lactococcus lactis
Plasmodium falcipa
Staphylococcus aur
Listeria monocytog
Listeria monocytog

```

XX PS Claim 2, Page 28-30; 32pp; Japanese.
XX CC The present invention describes a method for microbiologically producing
XX CC coenzyme Q10 at a high efficiency by using a gene of the synthesis of
XX CC coenzyme Q10 side chain originating in a fungus belonging to the genus
XX CC Saitoella. The present sequence represents the specifically claimed
XX CC Saitoella complicated protein having decaprenyl diphosphate synthase
XX CC activity. The protein having decaprenyl diphosphate synthase activity
XX CC can be used for producing coenzyme Q10.
XX CC
XX Sequence 440 AA;
SQ
Query Match 100.0%; Score 2194; DB 22; Length 440;
Best Local Similarity 100.0%; Pred. No. 2.9e-202;
Matches 440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASPARIRSSISSTISLSTVLTARTASAPRLRCPTSPSSMAAASASRLVEPD 60
DB 1 MASPARIRSSISSTISLSTVLTARTASAPRLRCPTSPSSMAAASASRLVEPD 60
QY 61 PNOPLINPLLVGPEMSNLTSNIRSLGSGHPSLDTVAKYVQSEGHIRPLVLMQAQ 120
DB 61 PNOPLINPLLVGPEMSNLTSNIRSLGSGHPSLDTVAKYVQSEGHIRPLVLMQAQ 120
QY 121 TEVAPKVOGMEKVEVPVNEGSLAPPEVLNDKNDPMNMNRSGPLTKGGEIRGQTSNIIASQ 180
DB 121 TEVAPKVOGMEKVEVPVNEGSLAPPEVLNDKNDPMNMNRSGPLTKGGEIRGQTSNIIASQ 180
QY 181 RLRLAETEMHHAASLHDDVIDASETRRNAPSGNQAFGNKALTAGPFLGRASVALARL 240
DB 181 RLRLAETEMHHAASLHDDVIDASETRRNAPSGNQAFGNKALTAGPFLGRASVALARL 240
QY 241 RNEVEIELLATVIANLVGEFEMQKNTVDDAIEATATQETFDVYLQKTYLKTASLIKSC 300
DB 241 RNEVEIELLATVIANLVGEFEMQKNTVDDAIEATATQETFDVYLQKTYLKTASLIKSC 300
QY 301 RASALLGATPEVADAAYAGRNIGLAFQIVDDMLDVTASATDIGRPAGADLQGLATAP 360
DB 301 RASALLGATPEVADAAYAGRNIGLAFQIVDDMLDVTASATDIGRPAGADLQGLATAP 360
QY 361 ALPAWKHAHELGPWKIKRFGDVERARELVEKSDGLEKTRALAEYAKALDAITFP 420
DB 361 ALPAWKHAHELGPWKIKRFGDVERARELVEKSDGLEKTRALAEYAKALDAITFP 420
QY 421 ESPARKALEQLTDKVLTRSR 440
DB 421 ESPARKALEQLTDKVLTRSR 440

RESULT 2
ABB83793
ID ABB83793 standard; Protein; 451 AA.
XX ABB83793;
XX
XX 29-AUG-2002 (first entry)
XX
XX Fungal decaprenyl diphosphate synthase SEQ ID NO 4.
XX
XX Fungi; Aspergillus; Leucosporidium; coenzyme Q10; enzyme;
XX decaprenyl diphosphate synthase; enzyme.
XX
XX Aspergillus clavatus.
XX
XX MO200252017-A1.
XX
XX 04-JUL-2002.
XX
XX 27-DEC-2001; 2001MO-JP11523.
XX
XX 27-DEC-2000; 2000JP-0398658.
XX

```

```

PA (KANF) KANEXA CORP.
XX
XX Matsuda H, Kawamukai M, Yajima K, Ikenaka Y;
XX
XX WPI; 2002-500767/53.
XX
XX N-PSDB; ABB81605.
XX
XX
XX Microbial production of coenzyme Q10 by transformants transferred with
XX PT fungal decaprenyl diphosphate synthase gene, on industrial scale for
XX PT application in drugs
XX
XX
XX Claim 4; Page 37-40; 52pp; Japanese.
XX
XX PS
XX CC The invention relates to a DNA sequence (I, ABB81604-ABB81605),
XX CC containing a sequence based on that of (I) but with some bases deleted,
XX CC added, inserted and/or substituted and encoding a protein
XX CC (ABB83792-ABB83793) with decaprenyl diphosphate synthase activity or
XX CC hybridisable with (I) under stringent conditions and encoding a protein
XX CC with decaprenyl diphosphate synthase activity. The DNA is useful for the
XX CC production of coenzyme Q10 for application in drugs. The coenzyme is
XX CC simply produced on industrial scale, efficiently, with significantly
XX CC productivity.
XX CC
XX Sequence 451 AA;
SQ
Query Match 56.7%; Score 1243; DB 23; Length 451;
Best Local Similarity 57.2%; Pred. No. 1.2e-110;
Matches 261; Conservative 61; Mismatches 108; Indels 26; Gaps 5;
QY 6 LRRTSSSSISLSTVLTARTASAPRLRCPTSPSSMAAASASRLVEPD 47
DB 1 MRRTVSSASGLT---LSRTTSTSTICWCUREDLNSNOVOIHVRKXHPTRKQVSPG 56
QY 48 AAVSASRLVEBDDPNOP--LINPLNVGPEMSNLTSNIRSLGSGHPSLDTVAKYVQ 104
DB 57 AAVSAAQITIFKGLPKAPPSISVDPLRIVEKELFPTKNIRQLGSGHPTLDVAKYTRRS 116
QY 105 EKGHIRPLVLMQAQTEVAPKVOGMEKVEVPVNEGSLAPPEVLNDKNDPMNMNRSGPLT 164
DB 117 EKGHMRPLVILMSQXTALTTPQSRSNFTPSQMVNDPISSPSVLADVTPLDPLVSKSAE 176
QY 165 KOGEIRGQTSNIIASORRLAETEMHHAASLHDDVIDASETRRNAPSGNQAFGNKAL 224
DB 177 AOYDFRAGD-ENTLPQORLAETELHTASLHDDVIDAIVARRSSNANLOFGKMAVL 235
QY 225 AGDFLLGRASVALARLNEVEIELLATVIANLVGEFEMQKNTVDDAIEATATQETFDVY 284
DB 236 AGDFLLGRASVALARLNEVEIELLATVIANLVGEFEMQKNTADDEKNPVTDDGTTISY 295
QY 285 LQKTYLKTASLIKSCRASALLGATPEVADAAYAGRNIGLAFQIVDDMLDVTASATDL 344
DB 296 LQKTYLKTASLIKSCRAAALLGSGTPEVADAAYAGRNIGLAFQIVDDMLDVTASATDL 355
QY 345 GKPAGADLQGLATAPALPAWKHAHELGPWKIKRFGDVERARELVEKSDGLEKTRAL 404
DB 356 GKPAGADLQGLATAPALPAWKONPPLGFLVGRKFSRBDGVMAKELVYKSDGVQETRAL 415
QY 405 ABEYAKALDAITRTPESPARKALEQLTDKVLTRSR 440
DB 416 ABEYAKALITAVSNFPDSEAKAGLQMCCKANRRK 451

RESULT 3
ABB83792
ID ABB83792 standard; Protein; 502 AA.
XX ABB83792;
XX
XX 29-AUG-2002 (first entry)
XX
XX Fungal decaprenyl diphosphate synthase SEQ ID NO 3.
XX
XX Fungi; Aspergillus; Leucosporidium; coenzyme Q10; enzyme;
XX

```


decaprenyl diphosphate synthase; enzyme.
Leucosporidium scottii.
WO200252017-A1.
04-JUL-2002.
27-DEC-2001; 2001WO-JP11523.
27-DEC-2000; 2000JP-03198658.
(KANF) KANEKA CORP.
Matsuda H, Kawamukai M, Yajima K, Ikenaka Y;
WPI; 2002-500767/53.
N-PSDB; ABN81604.
Microbial production of coenzyme Q10 by transformants transferred with
fungal decaprenyl diphosphate synthase gene, on industrial scale for
application in drugs -
Claim 3; Page 33-36; 52pp; Japanese.
The invention relates to a DNA sequence (I, ABN81604-ABN81605),
containing a sequence based on that of (I) but with some bases deleted,
added, inserted and/or substituted and encoding a protein
(AB83792-AB83793) with decaprenyl diphosphate synthase activity or
hybridisable with (I) under stringent conditions and encoding a protein
with decaprenyl diphosphate synthase activity. The DNA is useful for the
production of coenzyme Q10 for application in drugs. The coenzyme is
simply produced on industrial scale, efficiently, with significantly
productivity.
Sequence 502 AA;
Query Match 48.5%; Score 1064; DB 23; Length 502;
Best Local Similarity 47.8%; Pred. No. 2.4e-93;
Matches 238; Conservative 80; Mismatches 112; Indels 68; Gaps 11;
QY 6 LRIRSISSIASIRSVTLRTASAPSLRLCTPTSPSSSWAAV-----SSA 53
DB 10 LRGRAPSSLLQPTBLQKLSSTPSSLRHAPSR--SAWTAIPGLSATFPASTSTS 67
QY 54 SRLVEPDNPQPLINPLNLVGPENSLTNSIRSLGSGHPSLDTVAKYVQSEGHIRPLM 113
DB 68 SSLAGSSKVALQDPLKPLGAENGLLRNSVQHLLGSGHPALDTIAKYFFQAEKHYRPML 127
QY 114 VLLMAQATE-VAPKVGQWEK-----VVEV-----PVNEGLAPPEVLNDKNPD 154
DB 128 ILLMSQATNGLAP---GWEQRRDQAAAAELKREQDGLGGDDIDEPLSPSVLNDQNS 184
QY 155 MM-----NMRSGPLTKDGELEGQTSNILASORRLAEITEMIHAASLLHDDVID 202
DB 185 MLASAKSFSDPLASLRPATPTPSIAQSIHOTLLPSQRRLAEITEMIHAASLLHDDVID 244
QY 203 ASERRNAPSGNAGFKMMAILAGDFLLGRASVALARLRNPEVIELLATVIANLVEGEFM 262
DB 245 LAETRSAPSPSLFNGKLSLIGADDFLLARASLSLSLGSNEVELVASLVANLVEGEVM 304
QY 263 QLKNTV--DDAIBATA-----TOETPDYLLQKTYLKTASLIKSCRASALLGGA-- 309
DB 305 QMKGNVPGKEGLLAGAGGGSTAGKPTPEIFDHYMKTYLKTASLIKSTRATILGGCGV 364
QY 310 -----TPEVADAYAGRNGLGAFQVDDMLDYTSATDLGKP-AGADILGLATAPAL 362
DB 365 KQWAEKEGVKDYAYSVGRNGLGAFQVDDMLDFTASAAQVGRPGGADLKLGLATAPAL 424
QY 363 FAWKHAEELGPMIKRFPDQGVREARRELVEKSDGLEKTRALAEVEYAKALDAIRTPFES 422
DB 425 YANEEFPELGAMLERKPAGEDDVEQARHLISRSSGAERTALAEHSKLARQALEGLPDS 484

QY 423 PARKALEQLTDKULTRSR 440
DB 485 EARTALDNNMARDTLRSKK 502
RESULT 4
ID AAO14997 standard; Protein; 537 AA.
XX AAO14997;
XX 08-AUG-2002 (first entry)
XX Rhodotorula minuta decaprenyl diphosphate synthase 1.
XX Fungi; decaprenyl diphosphate synthase; enzyme; high-yield fermentation;
XX coenzyme Q10 production.
XX Rhodotorula minuta.
XX WO200240682-A1.
XX 23-MAY-2002.
XX 20-NOV-2001; 2001WO-JP10119.
XX 20-NOV-2000; 2000JP-0352940.
XX (KANF) KANEKA CORP.
XX Matsuda H, Kawamukai M, Yajima K, Ikenaka Y;
XX WPI; 2002-427097/45.
XX N-PSDB; AAL42940.
XX Decaprenyl diphosphate synthase gene of Rhodotorula origin for
XX efficient preparation of coenzyme Q10 -
XX Claim 3; Page 33-36; 50pp; Japanese.
XX The invention comprises the amino acid and coding sequences of fungal
XX (Rhodotorula minuta) decaprenyl diphosphate synthase enzymes. The fungal
XX decaprenyl diphosphate synthase DNA and protein sequences are useful for
XX the efficient, high-yield fermentative production of coenzyme Q10 - for
XX pharmaceutical use. The present amino acid sequence represents a
XX Rhodotorula minuta decaprenyl diphosphate synthase enzyme.
XX Sequence 537 AA;
Query Match 46.9%; Score 1028.5; DB 23; Length 537;
Best Local Similarity 48.6%; Pred. No. 6.9e-90;
Matches 237; Conservative 84; Mismatches 108; Indels 59; Gaps 12;
QY 2 ASPALIRISIRSIASIRSVTLRTASAPS-----LRLRCTPTSPSSSWAA 48
DB 60 SKPAARSTSTAPALPSPSTSDPQSSSPSSSSSSSILPDLFRPLSSSSSSSTSS 119
QY 49 AVSSASRLVEPDNPQPLINPLNLVGPENSLTNSIRSLGSGHPSLDTVAKYVQSEGH 108
DB 120 SSSSSSNRKNNTNSNTIFDPLQVLGNELSRLSNVQALLGSGHPALDTIAKYFFQAEK 179
QY 109 IRPLMYLLMAQATE-VAPKVGQWEKVEV-----PVNEGL--APPEVLNDKNP 153
DB 180 IRPMIVLLMSQATNGLAP---GFEERSKLELSGRKQTDPSRSINDPLEVKADSLINDSNP 236
QY 154 -DMNMRSGPLTKDGELEGQTSNILASORRLAEITEMIHAASLLHDDVIDASERRNAPS 212
DB 237 SSFAASSSSPL-----DSMFSTSNVLPFSQRRLAEITEMIHAASLLHDDVIDGSAMRRAQS 292
QY 213 GNQAFGNKMAILAGDFLLGRASVALARLRNPEVIELLATVIANLVEGEFMOLKNTVDDAI 272
DB 293 APAAFGNKISVLGGDFLLARASLYLSRLGSGNEVELVASLVANLVEGEVMQIKG---NAP 349

Oy	273	BATA-----	TOEFDYVYLKTYKTSIA	SLAKSRA	BALLGATPE-----	V	313
Db	350	ESNAGSGAEVAVHRLTP	LEFHEHYMKTYIKTSL	IAKSRAT	TIIGGAGECKGWIEGRI		409
Oy	314	ADAAVAYGNGLGAFQI	VDDMLDYVSATD	LCKPA-GAD	LQGLATPA	LPFWKHAEIG	372
Db	410	KDIASVSGNIGIAI	FQIVDDLDLFDATDA	QFGKPSQ	ADLTKGLATPA	LPALYAMEEFPENG	469
Oy	373	PIIKKFSFGPGEVARE	LVESDGLKTRALAE	YACALAI	ITPPSPARKALEQT		432
Db	470	QMIILKFFENEGVE	TPARNLVRSAGEP	EKTVKAEKHA	LALMALQGLP	PSDAREALEGLT	529
Oy	433	DKVLTFSR	440				
Db	530	KTVLNRTK	537				
RESULT 5							
ID	AA014998	standard; Protein; 425 AA.					
AC	AA014998;						
DT	08-AUG-2002	(first entry)					
DE	Rhodotorula minuta decaprenyl diphosphate synthase 2.						
XX	Fungi; decaprenyl diphosphate synthase; enzyme; high-yield fermentation;						
XX	coenzyme Q10 production.						
XX	Rhodotorula minuta.						
XX	WO200240682-A1.						
XX	23-MAY-2002.						
XX	20-NOV-2001; 2001WO-JP10119.						
XX	20-NOV-2000; 2000JP-0352940.						
XX	(KANF) KANEKA CORP.						
XX	Matsuda H, Kawamukai M, Yajima K, Ikenaka Y;						
XX	WPI; 2002-427097/45.						
XX	N-PSDB; AAL42941.						
XX	Decaprenyl diphosphate synthase gene of Rhodotorula origin for						
XX	efficient preparation of coenzyme Q10						
XX	Claim 4; Page 39-41; 50pp; Japanese.						
XX	The invention comprises the amino acid and coding sequences of fungal						
XX	(Rhodotorula minuta) decaprenyl diphosphate synthase enzymes. The fungal						
XX	decaprenyl diphosphate synthase DNA and protein sequences are useful for						
XX	the efficient, high-yield fermentative production of coenzyme Q10 - for						
XX	pharmaceutical use. The present amino acid sequence represents a						
XX	Rhodotorula minuta decaprenyl diphosphate synthase enzyme.						
XX	Sequence 425 AA;						
XX	Query Match 45.5%; Score 999; DB 23; Length 425;						
XX	Best Local Similarity 52.9%; Pred. No. 3.3e-87;						
XX	Matches 221; Conservative 66; Mismatches 85; Indels 46; Gaps 10;						
Oy	59	PDPNOLINPLNLVGP	EMNLTSTNTRSLGSGH	PSLDYAKTYV	QSGGHIRPLMVLMA	118	
Db	18	PNALNTIFPDLQLV	GNELSTLSRNSVQAL	IGSGHPALDI	TAKYFFQAEGRH	IRPMIVLLMS	77
Oy	119	GATE-VAPKVGQMEK	VVEV-----	PVNEGL--	APPEVLNDKVP-D	MMNMNRSGP	162
Db	78	QATNGIAP--	GFEEKSKLELSGRK	OTDPSRSTIND	LEVYKADEI	LINDSNPSSFA	SSSP 134

Oy	163	LTKGEIGGQSNIIASGRRLAEITEMIHAASLHDVYIDASETRRNAPSGOAFGNMA	222
Dd	135	L-----DSMPSTNSNVPSGRRLAEITEMIHVASLHDVDYDGSAMRQAASATAAGNKIS	190
Oy	223	ILAGDFLLGRASVALARLNPEVEILLATVIANIVESEFMOKNTVDALIEXTA-----	276
Dd	191	VLGGFLLARSLALYSIRLSGSENEVELVASIANIVEGEVMQIKG--NAPESNAGSKEV	247
Oy	277	-----TOSTPPLYOKTYLTAKSLIAKSCRASALLGGATPE-----VMDAAAYAGR N	323
Dd	248	AVHRIITPEIFEHYMKITYLTAKSLIAKSTRATTITLGGAGEKQGI EGERIKDIAYSGRN	307
Oy	324	LGLAQIVDDMLDYTVSATDLGKPA-GADLOGLATAPALPAMKHAEIGPWIKKFSDP	382
Dd	308	LGIATQLVLDLDDLPFATDAQGFKPSQGDALKGLATAPALVAMEEFPFGMGWILNKFENE	367
Oy	383	GDEVAREARLVESDGLKETRRALAEYAOALDALITPPESPARKALEQTDTCKYLTRSR	440
Dd	368	GDVEARRNLVKRSAGPEKTIVLAEKHALAMBALGGLPESDAREALFGLTKIVLNRTX	425
 RESULT 6 AAOI4999 ID AAOI4999 standard; Protein; 403 AA. XX AC XX AAOI4999; XX DT 08-AUG-2002 (first entry) XX DE Rhodotorula minuta decaprenyl diphosphate synthase 3. XX KW Fungi; decaprenyl diphosphate synthase; enzyme; high-yield fermentation; XX KM coenzyme Q10 production. XX OS Rhodotorula minuta. XX PN WO200240682-A1. XX PD 23-MAY-2002. XX PE 20-NOV-2001; 2001WO-JP10119. XX PR 20-NOV-2000; 2000JP-0352940. XX PA (KANF) KANEKA CORP. XX PI Matsuda H, Kawamukai M, Yajima K, Ikenaka Y; XX DR WPI; 2002-427097/45. XX DR N-PSTD; PAIL42942. XX PT Decaprenyl diphosphate synthase gene of Rhodotorula origin for XX PT efficient preparation of coenzyme Q10 - XX PS Claim 19; Page 44-46; 50pp; Japanese. XX CC The invention comprises the amino acid and coding sequences of fungal XX CC (Rhodotorula minuta) decaprenyl diphosphate synthase enzymes. The fungal XX CC decaprenyl diphosphate synthase DNA and protein sequences are useful for XX CC the efficient, high-yield fermentative production of coenzyme Q10 - for XX CC pharmaceutical use. The present amino acid sequence represents a XX CC Rhodotorula minuta decaprenyl diphosphate synthase enzyme. XX SQ Sequence 403 AA;			
Oy	65	LINPLNTVGPEKSNLTSNRSLSGSHPSLPTVAKYVQSBEKGIRIPMLVLMQAOTE-V	123
Dd	2	I FDPQLGVNGELSSLRSNVQALSGGHALDTIAKYFOABEKGIRIPMVLVMSQATNCL	61
 Query Match 45.4%; Score 996; DB 23; Length 403; Best Local Similarity 53.4%; Pred. No. 5.9e-87; Matches 220; Conservative 65; Mismatches 81; Indels 46; Gaps 10.			

CC present sequence represents a specifically claimed decaprenyl synthase.
 CC The method allows the side chain of ubiquinone to be freely modified
 CC by genetic engineering methods.

XX Sequence 377 AA;

Query Match 39.3%; Score 863; DB 18; Length 377;

Best Local Similarity 49.2%; Pred. No. 3,46-74;
 Matches 184; Conservative 61; Mismatches 101; Indels 28; Gaps 4;

XX 67 NPLNIVGPEMNLTSNIRSLGSHPSLDYVAKYVVOSEGHIRPLMLWLMQAQTEVAPK 126

DB 32 NASHIKNELQISRGIRQMLNSSEFLSECSKYTTIAQGMPSLVLMASKTSLSCHG 91

XX 127 VQGWKEVVEVNEGLAPPEVLANDKNPMNMNRSGPLTKDGEIGQTSNIIASQRIAEI 186

DB 92 ID-----RSVVGDKYIDDDLR-----FSTGQILPSQLRLAQI 125

XX 187 TEMTHAASLHDDVIDASETRRNAPSGNQAQGNKKAIIAGDFLLGRASVALARINPEVI 246

DB 126 TEMTHIASLHDDVIDANVRSGSSNVAFGNRSIIAGNFIARASTAMARLNPDVT 185

XX 247 ELATVIANLVGPEFMOLKNTVDDAIEATATQETFDYLYOKTYLKTASLIKSGRASAL 306

DB 186 ELATVIANLVGPEFMOLKNTVDDAIEATATQETFDYLYOKTYLKTASLIKSGRASAL 243

XX 307 GGATPEVADAAYAYGRNLGLAFOIVDDMLDYTSATDLGKPGADLOGLATAPALPAMK 366

DB 244 GQCSPSVATAGEYGRICGTAFQIAMDVDLYTSKDDTLGKAAGADLKLGATAPVLPAMK 303

XX 367 HHAELGPMIKKPSDPCGVERARELYEKSDELKTRALAEYAKALDAITFPESPARK 426

DB 304 KYPLGAMITVRFHMPSDIQRAGSLVECTDAIEQITWAKYIKKAKDSLCLPDSAPARK 363

XX 427 ALBQITDKVLTFRSR 440

DB 364 ALFALADKVTIRKK 377

XX Db

RESULT 9
 ABB58466 standard; Protein; 430 AA.

XX ID ABB58466;

XX AC ABB58466;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 2190.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX WO200171042-A2.

XX PN 27-SEP-2001.

XX PD 23-MAR-2001; 2001WO-US09231.

XX PF 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW,

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02569.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Disclosure; SEQ ID NO 2190; 21pp + Sequence Listing; English.

XX XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA

CC sequences (AB57737-AB572072).

CC (The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX XX

XX Sequence 430 AA;

Query Match 33.8%; Score 742; DB 22; Length 430;

Best Local Similarity 37.2%; Pred. No. 1,9e-62;

Matches 167; Conservative 85; Mismatches 135; Indels 62; Gaps 6;

XX 3 SPALRIRISIS-SRSIASLR---SVTLRTASAPSLRLRCTPSRPSSSVAAVSSASRLV 57

DB 33 APGYTINHIELETSKSSVRLQPSATKSTSYVQSPHKCSQSLQVSGSKANLRQHSVH 92

XX 58 EDDPNQPL---INPLNIVGPEMNLTSNIRSLGSG--HPSLDYVAKYVVOSEGHIRP 111

DB 93 TQOPAGPVREFOIDPYIILDDDLKTFYDDVRYLLKSGTSQPELDTIASYFPDQKALRP 152

XX 112 LNWLMQAQTEVAPKQGWKEVVEVNEGLAPPEVLANDKNDNMNMRSGPLTKDGEIEG 171

DB 153 MVTWLMKAKINYN-----HLLN 168

XX 172 QTSNIIASQRRLAETEMTHAASLHDDVIDASETRRNAPSGNQAQGNKKAIIAGDFLLG 231

DB 169 ESHQVLHKQRLALFSEWHSASLVHDDVIDQSDRRKGPSNALNMHKVMAGDYILS 228

XX 232 RASVALARLNPEVIELATVIANLVGPEFMOLKNTVDDAIEATATQETFDYLYOKTYL 291

DB 229 IASIMTARLRSDVTLVLSQILDLVQGEFMOLGSRETE-----NEFFAHYLTTRYRK 281

XX 292 TASLIKSGRASALGGATPEVADAAYAYGRNLGLAFOIVDDMLDYTSATDLGKPGAD 351

DB 282 TASLIANALKATVIAQADNVAEVAFOYGRNIGLAFOLVDDMLDFVSTBQMGKPTAAD 341

XX 352 LQGLATAPALPAMKHAEIGPMIKKPSDPCGVERARELYEKSDELKTRALAEYAK 411

DB 342 LKLGATAPVLPACEKYPPLNPMWRRFSEPCDVERAFELVHKSHGLDQTRFLAKKHNE 401

XX 412 ALDAITFPESPARKALEQITDKVLTFRSR 440

DB 402 AIRLAQELTESPYOKGLQVAVDLVINRMK 430

XX RESULT 10

XX AAB19412 standard; Protein; 321 AA.

XX ID AAB19412;

XX AC AAB19412;

XX DT 06-MAR-2001 (first entry)

XX DE A straight-chain class prenilyltransferase designated ATP18.

XX KW Prenilyltransferase; ATP1; ATP2; ATP3; ATP4; ATP5; ATP6; ATP7;
 KW ATP8; ATP9; ATP10; ATP11; ATP12; tocopherol; homogentisic acid;
 KW pnytylpyrophosphate; 2-methyl-6-pnytylbenzoquinol; antioxidant;
 KW nutritional supplement.

XX OS Arabidopsis sp.

XX PN WO200063391-A2.

XX AAU72775;
 AC 26-FEB-2002 (first entry)
 DT XX
 DE Arabidopsis tocopherol cyclase ATP8.
 XX
 KM Tocopherol cyclase; isoprenoid; cariant; cytosolic; noctropic;
 KM neuroprotective; ophthalmological; antiarthritic; anti-aging;
 KM antioxidant; food additive; seed oil; tocotrienol; free radical damage;
 KM cardiac diseases; cancer; cataract; retinopathy; Alzheimer's disease;
 KM neurodegeneration; arthritis; transgenic plant.
 XX Arabidopsis sp.
 OS
 PN MO200179472-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 13-APR-2001; 2001WO-US12334.
 XX
 PR 14-APR-2000; 2000US-0549848.
 PR 14-OCT-2000; 2000US-0688069.
 XX
 PA (MONS) MONSANTO TECHNOLOGY LLP.
 XX
 PI Subramaniam SS, Slater SC, Karberg K, Chen R, Valentin HE;
 PI Wong YH;
 DR N-PSDB; AAS96924.
 DR
 PT New nucleic acids encoding tocopherol cyclase, useful in increasing the
 PT production or modulation of tocopherol cyclase in a host plant,
 PT particularly those plants that produce vegetable oils for edible and
 PT industrial uses -
 PT
 XX Example 1; Page 113-114; 166pp; English.
 PS
 XX The invention relates to an isolated nucleic acid sequence encoding a
 CC tocopherol cyclase, an enzyme involved in the synthesis of
 CC tocopherol, an antioxidant isoprenoid. Also included are transgenic
 CC plants expressing tocopherol cyclase where the tocopherol is purified
 CC from the seed as a refined and deodorized oil produced by treating a
 CC tocopherol rich oil by distilling under low pressure and high
 CC temperature, where the refined oil has reduced free fatty acids and a
 CC substantial percentage of tocopherol present in the pretreated oil.
 CC The nucleic acid is useful in the increased production or
 CC modulation of tocopherol cyclase in a host plant or plant cell. The
 CC polynucleotides can be used as hybridisation probe for RNA, cDNA or
 CC genomic DNA to isolate full length DNAs or genomic clones encoding a
 CC polypeptide and to isolate cDNA or genomic clones of other genes having
 CC high sequence identity to the given polynucleotide sequences.
 CC Tocopherols and tocotrienols are useful as antioxidants and play an
 CC important role in protecting cells from free radical damage, in the
 CC prevention of cardiac diseases, cancer, cataract, retinopathy,
 CC Alzheimer's disease, neurodegeneration, and arthritis, and in
 CC anti-aging. The nucleic acid constructs are useful for the increased or
 CC decreased expression of tocopherol cyclase in plants involved in the
 CC production of vegetable oils for edible and industrial uses. The
 CC present sequence represents a tocopherol cyclase or tocopherol cyclase
 CC -like protein.
 CC
 XX
 SO Sequence 321 AA;
 Query Match 30.0%; Score 658; DB 23; Length 321;
 Best Local Similarity 42.0%; Pred. No. 1.5e-54;
 Matches 150; Conservative 64; Mismatches 93; Indels 50; Gaps 5;
 QY 92 PSJDTVAKYVVS--EGNHRPLWLLMAQATEVAPKVQGEKVEVAVNGLAPPEVLN 149
 DB 7 PKLASAAEFPPKRGVQKQFSTILLMAITLVN-----RVPPALI 47

QY 150 DKNPDMNMRSGPLTKDGEIEGOTSIIILASQRIAEITTEMIHAASLIHDVDVADSETRRN 209
 DB 48 GSTDIV-----TSELRVQRGIIEITETEMHVAASLIHDVDVADADTRRG 91
 QY 210 AFGNQAFGNKAILAGDFLLGRASVALARLNPEVIELATVIANLVEGEMQKNTVD 269
 DB 92 VGSILNVYMGKNSVLAGDFLLRACALALAKNTEVALLATRAVEHLVYGTMETTS--- 148
 QY 270 DATEATATGETPDYVYKQTYLKTASTIAKSCASALLGGATPEVADAAYAGRNGLAFQ 329
 DB 149 ---STEQRYSMDYVYQKTYVYASTIASNSCAVAVLNGQFAEVAVLAFEGYRNLGAFQ 204
 QY 330 IYNDMLDYVSAATDCKPRAGADLQGLATAPALPMKHAIEGPMIKKFSPPGVPEAR 389
 DB 205 LIDDILDFGTGSASLGKSLDIRHGVITAPLIFMEEPQRLREVVDVEKPPRVVIAL 264
 QY 390 ELVEKSDGLEKTRALAEYAKALDAIRFPESP-----ARKALEQITDKVLTSSR 440
 DB 265 EYLGKSKGIQRARELAMEHANLAAAIGSLPETDNEVYKRSRALIDLTRVITNKK 321
 RESULT 13
 ID ABP41763
 XX ABP41763 strand: Protein; 272 AA.
 AC
 XX ABP41763;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HHNG108, SEQ ID NO:2895.
 XX
 KM Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KM ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KM infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KM PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KM inflammatory condition; immune disorder; blood disorder;
 KM cardiovascular disorder; respiratory disorder; neurological disorder;
 KM gastrointestinal disorder; urinary system disorder; drug screening;
 KM gene therapy; chromosome mapping; forensic analysis;
 KM antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KM antiinflammatory; gynaecological; reproductive; chromosome 10p11.2-12.
 XX
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 PN
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US18569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 PI
 DR WPI; 2002-147878/19.
 DR N-PSDB; ABQ54840.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11; SEQ ID NO 2895; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,

treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct_sequences.

```

PR 15-APR-1999; 99US-0129899.
PR 30-JUL-1999; 99US-0146461.
XX
XX
PA (CALJ ) CALGENE LLC.
XX
XX
PI Savidge B, Lasner MW, Weiss JD, Post-Beittenmiller D;
XX
XX WPI; 2000-647519/62.
XX
XX
XX An isolated nucleic acid sequence encoding prenyltransferase used to
XX transform plant cells to increase the production of tocopherols -
XX
XX Example 1; Page 81; 114pp; English.
XX
XX The present sequence represents a prenyltransferase. The specification
XX describes prenyltransferases designated ATPT1, ATPT2, ATPT3, ATPT4,
XX CC ATPT5, ATPT6, ATPT7, ATPT8, ATPT9, ATPT10, ATPT11, and ATPT12. The
XX CC biosynthesis of alpha-tocopherol in higher plants involves the
XX CC condensation of homogentisic acid and phytylpyrophosphate to form
XX CC 2-methyl-6-phytylbenzoquinol, which can form various tocopherols. The
XX CC prenyltransferase polynucleotides are useful in transforming host cells
XX CC to alter the expression of prenyltransferase in these cells. The
XX CC transformed cells are used in the production of tocopherols which are
XX CC of use in the pharmaceutical industry as antioxidants and also in the
XX CC food industry as nutritional supplements.
XX
XX Sequence 297 AA;
XX
XX
XX Query Match 25.6%; Score 561; DB 21; Length 297;
XX Best Local Similarity 37.5%; Pred. No. 2.8e-45;
XX Matches 134; Conservative 61; Mismatches 88; Indels 74; Gaps 6;
XX
XX QY 92 PSLDTVAKYVQSS--EGGHIRPLMLVLLMAQATEVAPKVGQWKEVVPVNEGLAPPEVLN 149
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX DQ 7 PKLSAAEYFFKRGVQKGFSTILLMATALNV-----RVPEALI 47
XX
XX QY 150 DKNPDMNMRSGPITKQGETGQTSNLASQRRRLAEITEMIHAAASLLHDDVIDASERN 209
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX DQ 48 GSTDIV-----TSELVRQRGIAEITEMIHVASLLHDDVLDDADTRRG 91
XX
XX QY 210 APSGNQAFGNKMAILAGDFLLGRASVALARLNREVEILLATVIANLVEGEFMQLKTVTD 269
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX DQ 92 VGLSNLVVGNK-----VVALLATAVEHLVLTGETMEITS--- 124
XX
XX QY 270 DAIEATATQETFDYYLYOKYLTASIAKSCRASALLGGATPEVADAAYAGRNGLGAFQ 329
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX DQ 125 ----STEQRYSDYMQKYYKYTASLNSCKAVAVITGQTAEVAVLAPEYGRNGLGAFQ 180
XX
XX QY 330 IVDMDLDYVSATDLGPKAGADLOGLATAPALPAWKHAEALGPMIKRKFSDPDGVERAR 389
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX DQ 181 LDDILDPTGTSASLKGSLSDIRHGVTAPILFAMEEFPQLREVWPQVEKDPNRNVDIAL 240
XX
XX QY 390 ELVSGDGLKEXTRALAEYQAQKALDARTPESP-----ARKALEQITDKVLTRSR 440
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX DQ 241 EYLKSGKGIQIARELAWEHANLAAAGISLPETDNEDVKGRSRAILIDLTHRVITRNK 297
XX
XX
XX RESULT 15
XX ABB81704
XX ID ABB81704 standard; Protein; 297 AA.
XX
XX AC ABB81704;
XX
XX DT 16-AUG-2002 (first entry)
XX
XX DE Arabidopsis sp. ATPT1.
XX
XX KW Prenyltransferase; PTS; isoprenoid; Synecocystis; Arabidopsis; corn;
XX KW rice; wheat; leek; canola; cotton; tomato; biosynthetic flux;
XX KW tocopherols; ATPT1; polytransferase.
XX
XX OS Arabidopsis sp.
XX

```

PN MO200233060-A2.
 XX
 PD 25-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US42673.
 XX
 PR 14-OCT-2000; 2000US-0668071.
 XX
 PA (MONS) MONSANTO TECHNOLOGY LLC.
 XX
 PI Lasserer MM, Savidge B, Weiss JD, Mitsky TA, Post-Beltemmiller MA;
 PI Valentin HE;
 XX
 DR WPI; 2002-463312/49.
 XX
 XX Novel nucleic acid sequences encoding prenyltransferase derived from
 PT eukaryotic and prokaryotic sources useful for producing plants and
 PT seeds with altered tocopherol content and compositions -
 PS Claim 6; Page 110-111; 148pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid sequence encoding
 CC a prenyltransferase (PTS). The nucleic acid construct of the invention is
 CC useful for altering isoprenoid content and for producing an isoprenoid
 CC compound of interest in a host cell, such as *Synechocystis* sp., or a
 CC plant cell obtained from *Arabidopsis*, corn, rice, wheat, leek, canola,
 CC cotton or tomato. The construct is also useful for increasing
 CC biosynthetic flux in a host cell toward production of an isoprenoid
 CC compound such as tocopherol and tocotrienol. The polynucleotide is useful
 CC for producing plants and plant parts, such as seeds, enriched in
 CC tocopherols. The sequence shown represents *Arabidopsis* sp. ATPT1, a
 CC polytransferase identified in the invention.
 XX
 SQ Sequence 297 AA;
 Query Match 25.6%; Score 561; DB 23; Length 297;
 Best Local Similarity 37.5%; Pred. No. 2.8e-45;
 Matches 134; Conservative 61; Mismatches 88; Indels 74; Gaps 6;
 QY 92 PSIDTVAKYVQ--EGKHRLPLVLLMAQTEVAPKVGWEKVEVPVNEGAPPEVLN 149
 Db 7 PKLASAEYFFKRGVQKQFSTILLMATLVN-----RVPEALI 47
 QY 150 DKNPDMNMMSGFLTKGDELGGTNSILASQRLAEITEMIHAASLLHDYIDASETRRN 209
 Db 48 GESTDIV-----TSELRVQRGIAEITEMIHVASLHDDVDADTRRG 91
 QY 210 APSGNQAFGNMIALAGDFLLGRASVALARLNPEVIELATVIANLVEGPFMQLKNTVD 269
 Db 92 VGSINVMGNK-----VVALLATVAEHLVGTGETMEITS--- 124
 QY 270 DATEATATQETFDYLYLQTYLTKTASLIASCRASALGATPEVADAAYAGRNIGLAFQ 329
 Db 125 -----STGORYSMDYTMQTYTKTASLISNSCKAVAVLTGQTAEVAVLAFEGRNIGLAFQ 180
 QY 330 IVDMLDYVSATDLGKPAQADLQGLATAPALPAMKHAEELGPMIKRKSDDPGDEPAR 389
 Db 181 LIDIDILPFTGTSASIGKGSLSDIRHGVTTAPILFAMEEFPQLRVVDQVEKDPNNVDIAL 240
 QY 390 ELVSKSGLKTRALAEYAKALDAIRTPESP-----ARKALEQTLTDKVLTRSR 440
 Db 241 EYLKSKGIQKRELAMEHANLAAAIIGSLPETNEDVKRSRRALIDLTHRVITRNK 297

Search completed: January 16, 2003, 09:46:48
 Job time : 58 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 16, 2003, 09:46:36 ; Search time 24 Seconds
(without alignments)
539.420 Million cell updates/sec

Title: US-09-830-111A-2
Perfect score: 2194
Sequence: 1 MASPALRIRSISSRIASLR.....ESPAKALEQLDKVLTISR 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	445.5	20.3	333	4	US-09-025-819-29
2	445.5	20.3	333	4	US-09-808-126-29
3	445.5	20.3	333	4	US-09-803-951-29
4	439.5	20.0	335	4	US-09-673-018-2
5	409	18.6	320	3	US-09-217-609A-6
6	409	18.6	320	4	US-08-873-235B-6
7	397.5	18.1	325	3	US-09-009-895-2
8	346	15.8	357	4	US-09-134-001C-3250
9	336.5	15.3	325	3	US-09-217-609A-2
10	336.5	15.3	325	4	US-08-873-235B-2
11	283	12.9	393	3	US-09-187-050-28
12	280	12.8	393	3	US-09-187-050-27
13	278	12.7	393	3	US-09-187-050-2
14	278	12.7	393	3	US-09-187-050-14
15	278	12.7	393	3	US-09-187-050-16
16	278	12.7	393	3	US-09-187-050-18
17	278	12.7	393	3	US-09-187-050-20
18	278	12.7	393	3	US-09-187-050-22
19	278	12.7	393	3	US-09-187-050-24
20	278	12.7	393	3	US-09-187-050-26
21	278	12.7	393	3	US-09-187-050-29
22	278	12.7	393	3	US-09-187-050-30
23	278	12.7	393	3	US-09-187-050-31
24	278	12.7	393	3	US-09-187-050-32
25	278	12.7	393	3	US-09-187-050-33
26	274	12.5	393	3	US-09-187-050-34
27	264	12.0	285	3	US-09-187-050-12

28	260	11.9	330	1	US-08-410-167A-4	Sequence 4, Appli
29	260	11.9	330	2	US-08-898-560-1	Sequence 1, Appli
30	258	11.8	330	4	US-09-101-126-1	Sequence 1, Appli
31	252	11.5	377	4	US-09-420-211-2	Sequence 2, Appli
32	248	11.3	297	1	US-08-534-910B-7	Sequence 7, Appli
33	248	11.3	297	1	US-08-534-910B-8	Sequence 8, Appli
34	248	11.3	297	1	US-08-534-910B-10	Sequence 10, Appli
35	247	11.3	297	1	US-08-534-910B-6	Sequence 6, Appli
36	247	11.3	297	3	US-08-886-466-2	Sequence 2, Appli
37	247	11.3	297	4	US-09-475-304-2	Sequence 2, Appli
38	247	11.3	297	4	US-09-101-126-3	Sequence 3, Appli
39	247	11.3	297	4	US-09-367-528A-5	Sequence 5, Appli
40	243	11.1	297	4	US-09-367-528A-1	Sequence 1, Appli
41	241	11.0	297	4	US-09-367-528A-3	Sequence 3, Appli
42	239	10.9	291	4	US-09-275-742-2	Sequence 2, Appli
43	238	10.8	297	1	US-08-534-910B-9	Sequence 9, Appli
44	206	9.4	298	1	US-08-095-726-4	Sequence 4, Appli
45	206	9.4	298	1	US-08-096-043-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-025-819-29
; Sequence 29, Application US/09025819
; Patent No. 6225097
; GENERAL INFORMATION:
; APPLICANT: Obata, Shusei
; APPLICANT: Nishino, Tokuzo
; APPLICANT: Koyama, Tanetoshi
; APPLICANT: Sato, Yoshihiro
; TITLE OF INVENTION: DECAPRENYL DIPHOSPHATE SYNTHETASE GENE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KENYON & KENYON
; STREET: 1500 K Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,819
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 251675
; FILING DATE: 17-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Khalilian, Hourl
; REGISTRATION NUMBER: 39,546
; REFERENCE/DOCKET NUMBER: 10235/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-220-4200
; TELEFAX: 202-220-4201
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-025-819-29

Query Match 20.3%; Score 445.5; DB 4; Length 333;
Best Local Similarity 30.3%; Pred. No. 4.5e-35;
Matches 120; Conservative 64; Mismatches 125; Indels 87; Gaps 9;
QY 62 NQPLINPLNVGPE---MSNLTSNIRSLIGSGH-PSLDTVAKYVYQSECKHIRPLMLVL 116

```

Db 4 NENSVKPLDRISVELAGMDRVNLIIRRMASRHAPRIPEVTAHLVEAGGKRLPMLVL- 62
QY 117 MAQATEVAPKVGWEKVEVFNVEGLAPPEVLNDKNDPMNMNRSGPLTKDGEIEGQTSNI 176
Db 63 -----AAARLCGYQ-----GNSHV 76
QY 177 LASQRRLAETITMTHAASLHDDVIDASSTRRNAPSGNOAFGNKMAIAGPFLGRASVA 236
Db 77 L-----LAAAVEFIHTATLTHDDVDDESQGRGRPTANLMDNKSIVLVGDYLFARSFOL 131
QY 237 LARLRNPEVIELLATVIANLVGEFMOUKNTVDALIEATNOE---TFDYLOKTYLTKTA 293
Db 132 MADTESQVMWRILANASATIAEGEVLOL-----TAAQDVSTTEDTYIQIVRGKTA 181
QY 294 SLIAKSCRASALLGATPEVADAAYAGRNIGLAFQIVDDMLDYTSATDLGKPGADLO 353
Db 182 ALFSAATEAGAVAGADPAVQCALFDYGDALGIAFQIVDDLDYGGSTTTIGKXVGDPR 241
QY 354 LGLATAPALFA-----WKHHAELGPMIKRFSDPGDERARELVEKSDGLEKTR 402
Db 242 ERKLTLPVIAKAIARADAEERAFMER-----TIGQGRQDEADLATALIILRRREALLEAAR 295
QY 403 ALAEVYAKALDAIRTFESPARKALEQLTDKVLTR 438
Db 296 ADAIAMAGRAKALQAAPOPLRILADLADPVS 331

```

RESULT 2

```

US-09-808-126-29
Sequence 29, Application US/09808126
Patent No. 6410280

```

GENERAL INFORMATION:

```

APPLICANT: Obata, Shusei
Nishino, Tokuzo
Koyama, Tanetoshi
Sato, Yoshihiro

```

```

TITLE OF INVENTION: DECAPRENYL DIPHOSPHATE SYNTHETASE GENE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSSEE: KENYON & KENYON
STREET: 1500 K Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/808,126
FILING DATE: 08-May-2001
CLASSIFICATION: <Unknown>

```

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 09/025,819
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Khalilian, Hourti
REGISTRATION NUMBER: 39,546
REFERENCE/DOCKET NUMBER: 10235/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-220-4200
TELEFAX: 202-220-4201

```

INFORMATION FOR SEQ ID NO: 29:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 29:

```

```

US-09-808-126-29

```

```

Query Match 20.3%; Score 445.5; DB 4; Length 333;
Best Local Similarity 30.3%; Pred. No. 4.5e-35;
Matches 120; Conservative 64; Mismatches 125; Indels 87; Gaps 9;
QY 62 NOPLINPLNVGE-----MSNTSINRSLGSGH-PSLDYAKYVYQSEGKIRPLMVL 116
Db 4 NENSVKPLDRISVELAGMDRVNLIIRRMASRHAPRIPEVTAHLVEAGGKRLPMLVL- 62
QY 117 MAQATEVAPKVGWEKVEVFNVEGLAPPEVLNDKNDPMNMNRSGPLTKDGEIEGQTSNI 176
Db 63 -----AAARLCGYQ-----GNSHV 76
QY 177 LASQRRLAETITMTHAASLHDDVIDASSTRRNAPSGNOAFGNKMAIAGPFLGRASVA 236
Db 77 L-----LAAAVEFIHTATLTHDDVDDESQGRGRPTANLMDNKSIVLVGDYLFARSFOL 131
QY 237 LARLRNPEVIELLATVIANLVGEFMOUKNTVDALIEATNOE---TFDYLOKTYLTKTA 293
Db 132 MADTESQVMWRILANASATIAEGEVLOL-----TAAQDVSTTEDTYIQIVRGKTA 181
QY 294 SLIAKSCRASALLGATPEVADAAYAGRNIGLAFQIVDDMLDYTSATDLGKPGADLO 353
Db 182 ALFSAATEAGAVAGADPAVQCALFDYGDALGIAFQIVDDLDYGGSTTTIGKXVGDPR 241
QY 354 LGLATAPALFA-----WKHHAELGPMIKRFSDPGDERARELVEKSDGLEKTR 402
Db 242 ERKLTLPVIAKAIARADAEERAFMER-----TIGQGRQDEADLATALIILRRREALLEAAR 295
QY 403 ALAEVYAKALDAIRTFESPARKALEQLTDKVLTR 438
Db 296 ADAIAMAGRAKALQAAPOPLRILADLADPVS 331

```

RESULT 3

```

US-09-803-951-29
Sequence 29, Application US/09803951
Patent No. 6413761

```

GENERAL INFORMATION:

```

APPLICANT: Obata, Shusei
Nishino, Tokuzo
Koyama, Tanetoshi
Sato, Yoshihiro

```

```

TITLE OF INVENTION: DECAPRENYL DIPHOSPHATE SYNTHETASE GENE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSSEE: KENYON & KENYON
STREET: 1500 K Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/803,951
FILING DATE: 13-Mar-2001
CLASSIFICATION: <Unknown>

```

PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 09/025,819
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Khalilian, Hourti
REGISTRATION NUMBER: 39,546
REFERENCE/DOCKET NUMBER: 10235/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-220-4200
TELEFAX: 202-220-4201

```

INFORMATION FOR SEQ ID NO: 29:

```

SEQUENCE CHARACTERISTICS:

```



```

Db 61 AC-----GNTN--LKH 69
QY 180 ORLAETEMIMHAASLHDDVIDASSTRNAPSGNQAFGNKMAILAGDFLLGRASVALAR 239
Db 70 AOKLAIIEMHTATLVHDDVDDESGLRRGPTANATWNNQTAVALVGVDFLIARAFOLLVD 129
QY 240 LRNPEVELLATVIANIVEGEFMOLKNTVDDAIEATATQETFDYLOKTYLKTASLIAS 299
Db 130 LDNMILLKDFSTGCTAEAGEVQLQ-----AQHPPTTE--DIYLOQIHGKTSRUFELA 182
QY 300 CRASALLGGATPEVADAAAYAGRNGLIAGFOIVDDMLDVTYSATDLGKPGAGADLQGLATA 359
Db 183 TEGAAIILAG-KPEVREPLRFAGFGNAFOIIDDILDTSDATLGNIGDGLMEGKPTL 241
QY 360 PALPANKH-HAELGPMIKRKFSPG--DVERARELVEKSGLEKTRALABEYAKALDAI 416
Db 242 FLIAAMQNTQGEORDLIRRSIATGTSQLEQVIAIVQNSGALDYCHKRATEETERALQAL 301
QY 417 RTPPSPARKALQDLTKVLTR 438
Db 302 EILPESTYROALVNLRLALDR 323

RESULT 8
US-09-134-001C-3250
; Sequence 3250, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3250
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3250

Query Match 15.8%; Score 346; DB 4; Length 357;
Best Local Similarity 24.8%; Pred. No. 2.4e-25;
Matches 95; Conservative 73; Mismatches 147; Indels 68; Gaps 6;

QY 65 LINPLNL-VGPENSLTNSRLLSGHPSLDTVAKYVVOSEGHKIRPLMLVLLMAQATEV 123
Db 36 VINVAKLININNEIKKVEKLEAEIISDQTLQEASFHLLSSGGKRVPAFVILSGQF--- 92
QY 124 APKVQGEKVEVPVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNILASORRL 193
Db 93 -----GSNNKPSEDTYR 105
QY 184 AEITEMIHAASLHDDVIDASSTRNAPSGNQAFGNKMAILAGDFLLGRASVALARLNP 243
Db 106 AVALLEIHMATLVHDDVIDSKRGRGLTISKWDQSTALTGNFLLAMGLKHLSEISDT 165
QY 244 EVIELLATVIANLVEGEFMOLKNTVDDAIEATATQETFDYLOKTYLKTASLIASCRAS 303
Db 166 RVHSTISKSIVDCRGELFQFQDN-----SNQITNVLRRNRKNTALLQLSTQVG 218
QY 304 ALLGGATPEVADAAAYAGRNGLIAGFOIVDDMLDVTYSATDLGKPGAGADLQGLATAPALF 363
Db 219 AITSNASNDVIRLKMTHGVHGSFQIIDDVLDFTSSEKKLGRKPVGSDLMNCHITLPVLL 278
QY 364 AKHHAELGPMIKRKF--DPGVERARE---LVEKSDGLEKTRALABEYAKALDAIR 417
Db 279 EMRKN-----KTFKDKISQLNPDSPQAFETCITIIROSESIBOSKQISEKILNKAINLID 334

```

```

QY 418 TTPPSPARKALQDLTKVLTRSR 440
Db 335 ELEDGPNKELFRKLKXGSRNK 357

RESULT 9
US-09-217-609A-2
; Sequence 2, Application US/09217609A
; Patent No. 6071733
; GENERAL INFORMATION:
; APPLICANT: MURAMATSU, Masayoshi
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OGURA, Kyozi
; APPLICANT: KOYAMA, Tanetoshi
; APPLICANT: SHIMIZU, Naoto
; APPLICANT: CHO, Yenwin
; TITLE OF INVENTION: Prenyl Diphosphate Synthetase Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, NW - Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/217,609A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/873,235
; FILING DATE: 11-Jun-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: TOFFENETTI, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 10235/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-217-609A-2

Query Match 15.3%; Score 336.5; DB 3; Length 325;
Best Local Similarity 26.7%; Pred. No. 1.7e-24;
Matches 100; Conservative 59; Mismatches 148; Indels 67; Gaps 6;

QY 72 VGEMSNLTNSRLLSGHPSLDTVAKYVVOSEGHKIRPLMLVLLMAQATEVAPKVQGW 131
Db 10 LNPYIIEVEKRLYECIQSDSEINAKAHILSSGGKRVPMFVLLSG----- 56
QY 132 KVEVPVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNILASORRLAEITEMIH 191
Db 57 -----FLNDTQKDDL-----IRTAVSLELVH 77
QY 192 AASLLHDDVIDASSTRNAPSGNQAFGNKMAILAGDFLLGRASVALARLNPVEIELLAT 251
Db 78 MASLVHDDVIDNSDMRGNTSVHIAFDKDTAIRTGHLLARALQNTATINNSKFQIFSK 137
QY 252 VIANLVEGEFMOLKNTVDDAIEATATQETFDYLOKTYLKTASLIASCRASALLGGATP 311
Db 138 TILVCFGEFDQWADFNVPVSFTA-----YLRINKRTAILIEASCHLGLSOLDE 190
QY 312 EVADAAAYAGRNGLIAGFOIVDDMLDVTYSATDLGKPGAGADLQGLATAPALFAWKHAE- 370

```


Db 276 IH-----KTAVLLECSVSGGILGGATEDEIARIIRYARCVGLLFQVDDILDVTKSSE 329
 QY 343 DLGKPGADLQGLATAPALFAWKHHAELGPMIKRKFSDFDVERARELVKESDGLKTR 402
 Db 330 ELGKTAGKDLTDTKATPKLM-----GLEKAK 356
 QY 403 ALAEEYAQAALDAIRTPFESPARKALEQLTDKVLTR 438
 Db 357 EFAELATRAKEELSSFDQIKAAPLL-GLADYIAFR 391

RESULT 12
 US-09-187-050-27
 ; Sequence 27, Application US/09187050B
 ; Patent No. 6043072
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney B
 ; APPLICANT: Hefner, Jerry
 ; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
 ; TITLE OF INVENTION: Synthese, And Methods of Use
 ; FILE REFERENCE: WSUR12423
 ; CURRENT APPLICATION NUMBER: US/09/187,050B
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 27
 ; LENGTH: 393
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
 ; OTHER INFORMATION: synthase protein variant
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)..(393)
 ; OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
 ; OTHER INFORMATION: variant
 US-09-187-050-27

Query Match 12.8%; Score 280; DB 3; Length 393;
 Best Local Similarity 25.9%; Pred. No. 7.2e-19;
 Matches 118; Conservative 59; Mismatches 171; Indels 108; Gaps 15;

QY 23 TLRTASAPSLRLCTPTSRSSSWAAVSSASRLVEPDNPQINPLNVLGPEMSNITSN 82
 Db 4 TAAAGTQSLQLRTVASYQCSNMRSCFK-----LTPFK-----SPHGWNFN 45
 QY 83 IRSILGS-----GHPSLDTYAKYVQSEGKHIRPLVLLMAQATEVAPKVGKVEVP 137
 Db 46 VPSLAANCAIMGHKLGLSLPYKQCSVSSKSTK-----TWAQLVDLAETKAEGKOIEFD 100
 QY 138 VNEGLAPPEVLNDKNDPMNMNRSGPLTKDGEI-EGQTSNIIASORRL----- 183
 Db 101 FNEYMKS KAVAVDAALD---KAIPLEYPEKIHESMRYSLLAGGKVRPAPALCIAACELVG 156
 QY 184 -----AEITEMIHAASLLHDDV--IDASETRRNAPSGNQAFGNKMAILAGDFLLGR 232
 Db 157 GSQDLAMPTACAMEMIHTMSLIHDDLPCMDNDPFRGKPTNKHVFGEDTAVLAGDALLSF 216
 QY 233 A----SVALARL-----RNPEVIELLATVIAN--LVEGEFMQLKNTVDIATEATQETFD 282
 Db 217 AFBIATVATKTVPSDRTLRISELGKTIGSQGLVGQGVDDITSEGDNVD-LKTLWEIH 275
 QY 283 YYLQKTYLTASLIASCRASALIGGATPEVADAAYAGNGLAFOIVDDMLDITYSAT 342
 Db 276 IH-----KTAVLLECSVSGGILGGATEDEIARIIRYARCVGLLFQVDDILDVTKSSE 329
 QY 343 DLGKPGADLQGLATAPALFAWKHHAELGPMIKRKFSDFDVERARELVKESDGLKTR 402
 Db 330 ELGKTAGKDLTDTKATPKLM-----GLEKAK 356
 QY 403 ALAEEYAQAALDAIRTPFESPARKALEQLTDKVLTR 438

Db 357 EFAELATRAKEELSSFDQIKAAPLL-GLADYIAFR 391

RESULT 13
 US-09-187-050-2
 ; Sequence 2, Application US/09187050B
 ; Patent No. 6043072
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney B
 ; APPLICANT: Hefner, Jerry
 ; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
 ; TITLE OF INVENTION: Synthese, And Methods of Use
 ; FILE REFERENCE: WSUR12423
 ; CURRENT APPLICATION NUMBER: US/09/187,050B
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 393
 ; TYPE: PRT
 ; ORGANISM: Taxus canadensis
 US-09-187-050-2

Query Match 12.7%; Score 278; DB 3; Length 393;
 Best Local Similarity 28.9%; Pred. No. 1.1e-18;
 Matches 103; Conservative 47; Mismatches 127; Indels 80; Gaps 11;

QY 117 MAQATEVAPKVGKVEVPVNEGLAPPEVLNDKNDPMNMNRSGPLTKDGEI-EGQTSN 175
 Db 80 MAQLVDLAETKAEGKIDIEFDNFNEYMKS KAVAVDAALD---KAIPLEYPEKIHESMRYSL 135
 QY 176 ILASORRL-----AEITEMIHAASLLHDDV--IDASETRRNAP 211
 Db 136 LLAGGKVRPAPALCIAACELVSGSQDLAMPTACAMEMIHTMSLIHDDLPCMDNDPFRGK 195
 QY 212 SGNAQGNKMAILAGDFLLGR-----SVALARL-----RNPEVIELLATVIAN--LVEGEF 261
 Db 196 TNHKVFGEDTAVLAGDALLSFATVATKTVPSDRTLRISELGKTIGSQGLVGQGV 255
 QY 262 MOLKNTVDDAIEATATQETFDYVLYQTYLTKLASIKRASCASALLGGATPEVADAAYAG 321
 Db 256 VDIITSEGDNVD-LKTLWEIH-----KTAVLLECSVSGGILGGATEDEIARIIRYA 308
 QY 322 RNGLAFQIVDDMLDITYSATDILGKPGADLQGLATAPALFAWKHHAELGPMIKRKFS 381
 Db 309 RCVGLLFQVDDILDVTKSSELGKTAGKDLTDTKATPKLM----- 350
 QY 382 PGDVERARELVKESDGLKTRALAEYQAQALDAIRTPFESPARKALEQLTDKVLTR 438
 Db 351 -----GLEKAKFAELATRAKEELSSFDQIKAAPLL-GLADYIAFR 391

RESULT 14
 US-09-187-050-14
 ; Sequence 14, Application US/09187050B
 ; Patent No. 6043072
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney B
 ; APPLICANT: Hefner, Jerry
 ; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
 ; TITLE OF INVENTION: Synthese, And Methods of Use
 ; FILE REFERENCE: WSUR12423
 ; CURRENT APPLICATION NUMBER: US/09/187,050B
 ; CURRENT FILING DATE: 1998-11-05
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 14
 ; LENGTH: 393
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
 ; OTHER INFORMATION: synthase protein variant

FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-14

Query Match 12.7%; Score 278; DB 3; Length 393;
Best Local Similarity 28.9%; Pred. No. 1,1e-18;
Matches 103; Conservative 47; Mismatches 127; Indels 80; Gaps 11;

QY 117 MAQTEVAPKVGMEKVEVPVNEGALPPEVLNDKPNPMNMRSGPLTKDGEI-EGQTSN 175
DB 80 MAQVLDLAETEKAGKDIPEFNEYMKSAAVADAALD---KAIPLLEYPEKIHESMRYS 135
QY 176 ILASQRL-----AETEMITHAASLHDV--IDASETRRNAP 211
DB 136 LLAGKRVPRALCIAACELVGGSDLAMPTACAMEMHTMSLIHDDLPCMDNDPFRGKP 195
QY 212 SGNQAFGNKAILAGDFLLGRA---SVALARL---RNPEVIELLATVIAN--LVGEF 261
DB 196 TNHKVFGEPTAVLAGDALLSFAPFHHIAVATSKVPSDRTLRVISELKTIGSQGLVGGQV 255
QY 262 MQLKNTVDDAIEATATQETFDYVYLOKTYLKTASLIKSCRSALLGGATPEVADAAYAG 321
DB 256 VDTSEGDANVD-LKTLIEWIH-----KTAVLLECSVSGILGGATDEIARIRRYA 308
QY 322 RNGLAFOIVDDMLDYVSATDLGKPAADLQGLATAPALFAWKHHAEELGPMIKRKFSD 381
DB 309 RCVGILLFQVVDILDYTKSSSEELGTAGKDLITDKATYPKLM-----350
QY 382 PGDVERARELVEKSDGLEKTRALAEVAKALDAIRTPESPARKALEOLTQKVLTR 438
DB 351 -----GLEKAKEFAAEELATRAKEELSSFDQIKAAPLL-GLADYIAFR 391

RESULT 15
US-09-187-050-16
Sequence 16, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Hefner, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranyl Diphosphate
FILE REFERENCE: MSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 16
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-16

Query Match 12.7%; Score 278; DB 3; Length 393;
Best Local Similarity 28.9%; Pred. No. 1,1e-18;
Matches 103; Conservative 47; Mismatches 127; Indels 80; Gaps 11;

QY 117 MAQTEVAPKVGMEKVEVPVNEGALPPEVLNDKPNPMNMRSGPLTKDGEI-EGQTSN 175
DB 80 MAQVLDLAETEKAGKDIPEFNEYMKSAAVADAALD---KAIPLLEYPEKIHESMRYS 135
QY 176 ILASQRL-----AETEMITHAASLHDV--IDASETRRNAP 211

DB 136 LLAGKRVPRALCIAACELVGGSDLAMPTACAMEMHTMSLIHDDLPCMDNDPFRGKP 195
QY 212 SGNQAFGNKAILAGDFLLGRA---SVALARL---RNPEVIELLATVIAN--LVGEF 261
DB 196 TNHKVFGEPTAVLAGDALLSFAPFHHIAVATSKVPSDRTLRVISELKTIGSQGLVGGQV 255
QY 262 MQLKNTVDDAIEATATQETFDYVYLOKTYLKTASLIKSCRSALLGGATPEVADAAYAG 321
DB 256 VDTSEGDANVD-LKTLIEWIH-----KTAVLLECSVSGILGGATDEIARIRRYA 308
QY 322 RNGLAFOIVDDMLDYVSATDLGKPAADLQGLATAPALFAWKHHAEELGPMIKRKFSD 381
DB 309 RCVGILLFQVVDILDYTKSSSEELGTAGKDLITDKATYPKLM-----350
QY 382 PGDVERARELVEKSDGLEKTRALAEVAKALDAIRTPESPARKALEOLTQKVLTR 438
DB 351 -----GLEKAKEFAAEELATRAKEELSSFDQIKAAPLL-GLADYIAFR 391

Search completed: January 16, 2003, 09:50:45
Job time : 25 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 16, 2003, 09:49:07 ; Search time 19 Seconds
(without alignments)
460.345 Million cell updates/sec

Title: US-09-830-111a-2
Perfect score: 2194
Sequence: 1 MASPALRBSISRSIASLR.....ESPAKALQLDKVLTISR 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	451.5	20.6	323	10	US-09-815-242-10321
2	451.5	20.6	323	10	US-09-815-242-10321
3	445.5	20.3	329	10	US-09-815-242-11142
4	405	18.5	322	10	US-09-815-242-12031
5	371.5	16.9	312	10	US-09-815-242-12603
6	367.5	16.8	312	10	US-09-815-242-5315
7	347	15.8	326	10	US-09-815-242-10573
8	340	15.5	350	9	US-09-925-637-66
9	339	15.5	273	10	US-09-815-242-11306
10	302.5	13.8	307	10	US-09-815-242-10069
11	272	12.4	367	9	US-10-108-915-26
12	260	11.9	299	10	US-09-815-242-1069
13	255.5	11.6	350	9	US-10-108-915-22
14	254.5	11.5	232	9	US-10-108-915-14
15	253	11.5	369	9	US-10-108-915-18
16	252	11.5	377	10	US-09-934-778-2
17	251.5	11.5	293	10	US-09-815-242-10630
18	247.5	11.3	316	9	US-10-108-915-45
19	247	11.3	371	9	US-09-738-626-5880

20	245.5	11.2	291	10	US-09-815-242-13273	Sequence 13273, A
21	241	11.0	295	10	US-09-815-242-11971	Sequence 11971, A
22	240	10.9	291	10	US-09-815-242-13597	Sequence 13597, A
23	237.5	10.8	299	10	US-09-815-242-14084	Sequence 14084, A
24	235.5	10.7	228	9	US-10-108-915-12	Sequence 12, Appl
25	226.5	10.3	345	9	US-10-108-915-44	Sequence 44, Appl
26	212	9.7	342	9	US-10-108-915-34	Sequence 34, Appl
27	202.5	9.2	209	9	US-10-108-915-30	Sequence 30, Appl
28	193	8.8	303	9	US-09-941-947A-26	Sequence 26, Appl
29	191.5	8.7	295	10	US-09-815-242-11239	Sequence 11239, A
30	189	8.6	287	10	US-09-925-637-64	Sequence 64, Appl
31	189	8.6	293	10	US-09-815-242-12583	Sequence 12583, A
32	187.5	8.5	297	9	US-09-941-947A-20	Sequence 20, Appl
33	187.5	8.5	297	10	US-09-934-903-14	Sequence 14, Appl
34	187.5	8.5	297	10	US-09-934-868-72	Sequence 72, Appl
35	185.5	8.5	382	9	US-09-738-626-4197	Sequence 4197, Ap
36	184.5	8.4	288	10	US-09-815-242-5239	Sequence 5239, Ap
37	182	8.3	326	9	US-10-108-915-46	Sequence 46, Appl
38	180.5	8.2	342	9	US-10-108-915-38	Sequence 38, Appl
39	179.5	8.2	133	9	US-10-108-915-4	Sequence 4, Appl
40	170	7.7	303	10	US-09-815-242-11382	Sequence 28, Appl
41	164.5	7.5	128	9	US-10-108-915-28	Sequence 1, Appl
42	158	7.2	295	10	US-09-547-267-1	Sequence 11547, A
43	156	7.1	303	10	US-09-815-242-11547	Sequence 614, App
44	142.5	6.5	304	10	US-09-925-302-614	Sequence 16, Appl
45	141	6.4	316	9	US-10-108-915-16	

ALIGNMENTS

RESULT 1
US-09-815-242-10321
; Sequence 10321, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10321
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10321

Query Match 20.6%; Score 451.5; DB 10; Length 323;
Best Local Similarity 31.0%; Pred. No. 1.2e-30;

Matches	119;	Conservative	49;	Mismatches	133;	Indels	65;	Gaps	6;
Qy	71	LVGPEMSNLTSNIRSLGSGHPSLDTVAKYVYVQSGGKHIRPLMWLLMQAATEVAPKVQGW	130						
Db	15	LADPMQVQNQVILAQNLSDVPLICQLGFYIVQGGGKHIRPLIAVLAARSL-----	65						
Qy	131	EKWVEVPYNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNILASQRRLABITEMI	190						
Db	66	-----GFEGNSITCAT-----FVBEFI	82						
Qy	191	HAASLLHDDVDVDSATRRNPASGNOAFGNKMAILAGDFLLGRASVALARLNPEVIELLA	250						
Db	83	HTASLLHDDVDDESMDRGRATANAEEFGNAASVLVGDFTYTFAPQLVAQLESKLILSIWA	142						
Qy	251	TVIANLVGEERFQMKNVVD-DAIEATATAQETFDYVYLOKTYLKTASLIAKSCRASALLGGA	309						
Db	143	DATNVLAESGEVQQLMVNDPSETSEAN-----YMRVIYKTARLFVAGQAAIVAGG	194						
Qy	310	TPEVADAAYAYGRNIGLAFQIVDDMLDVTYSATDGLGKPADAGDLQGLGATAPALFAWKH-H	368						
Db	195	TEAQEKALQDYGRLGTAFQVLDVDDVSANTQALGKNVGDGLAEGKPTLPLHAMRHGN	254						
Qy	369	AEIGPMIYKRKPSDPGDVERARE---LVPEKSDGLEKTRALAEYAAQKALDAITPTFESPAP	425						
Db	255	AQQAALIREATIEQGGKREAIDEVLAIMTEHKSLDYAMNRAKEEAQKAVDAIEILPESEYK	314						
Qy	426	KALEQL	431						
Db	315	QALISL	320						

```

RESULT 4
US-09-815-242-12031
; Sequence 12031, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L. W.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12031
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12031

```

Query Match	18.5%;	Score 405;	DB 10;	Length 322;
Best Local Similarity	29.8%;	Pred. No. 9.8e-27;		

Matches	111;	Conservative	59;	Mismatches	140;	Indels	62;	Gaps
Qy	71	LVGPEMNLTSNRISLGGHPSLDTVAKYVVOSEGHKHIRPLMLVLLMAQAATEVPKVQGW	130					
Db	9	VVAEDFTAVDGIIRQLTSTRVPLVEKIGDYIIISAGGKRRLRPLVLVLL-----AGKTGY	61					
Qy	131	EKKVEVPVNEGLAPPELVNLDKPMMNMRSGLPTKDGEIEGOTSNILASORRLAEITEMI	190					
Db	62	-----KGDDLC-----LLAATIEFL	76					
Qy	191	HAASLLHHDDVIDASESTRRNAPSNGNAFGNKMAILAGDFLLGRASVALARLNREVIELLA	250					
Db	77	HTSILLHDDVVDAAGSLRRGRSTANALWGNAPSVLGVDFLYARSFEMMVELGSPVMRIIS	136					
Qy	251	TIVIANLVEFMQJKNVTVDPAIEATATOETFDYYLKTYLKTASLIAKSCRASALICGAT	310					
Db	137	QATRVIAEGEVLQLSKVRD---ASTTEET--YMEVIRGKTAWLFEASTHSAAALCQAG	189					
Qy	311	PEVADAAYAGRNIUGLAFQIVDDMLDTVTSGADTLGKRPAGADLOLGUATAPALFAWKHAE	370					
Db	190	EEQSEALRRFGDYLGIAFQVLVDLLDYRGDAATLGNKVGDDLAE GKPTLP LITMRDGTE	249					
Qy	371	LGPWKIKKFSDPG---DVERARELUVEKSDGLEKTRALAEYAAOKALDAIRTFPESPARKA	427					
Db	250	EQAALVRKAIOOGGSQDLSECAAVEAAGALDYTDANLARDYAARAACLTLPDNEYRSA	309					
Qy	428	LEQLTDCVKLTRS	439					
Db	310	LVELSEFAVART	321					

```

RESULT 5
US-09-815-242-12603
; Sequence 12603, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L. W.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12603
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12603

```

Query Match 16.9%; Score 371.5; DB 10; Length 322;
Best Local Similarity 26.8%; Pred. No. 6.6e-24;

	Matches	101;	Conservative	69;	Mismatches	144;	Indels	63;	Gaps	57.
QY	69	LNLVGPENSLTNSRSLSGSPISLDTVAK----	YYVGSEGGHIRPLMLVLLMAQAATEYA	124	:	:	:	:	:	:
Dd	2	INVALTNMNNELKKYEGRLEKAIKSDSVLEQASHLSSGGRAPPAVLISSQ-----		56	:	:	:	:	:	:
QY	125	PKVGOMEKVVEPVNEG LAPREV LNDKNDDMMNBSGLTKDGEIEGOTSNIILASORILA	184	:	:	:	:	:	:	:
Dd	57	-----FGKDGEQSSEQTYGV-----A	71	:	:	:	:	:	:	:
QY	185	EITEMITHAASLIHDVIDASETRRNAPSNGOAFGNKMAILAGDFLLGRASVALLARLNDE	244	:	:	:	:	:	:	:
Dd	72	VALLFIHMATLVHDDVIDIDSKDKRGKRGLTIKSKMDQTALTITGNFLALGLEHLMVKNR	131	:	:	:	:	:	:	:
QY	245	VIELLATIYANIUVEGEFMOLKNTVDALERTATQTEFFDYILOKYUKTASLIAKSCRASA	304	:	:	:	:	:	:	:
Dd	132	VHQLISESIVDCRGELTFEQDOFN-----SQGTIIINYLRIRNRKTALLIQISTEYGA	184	:	:	:	:	:	:	:
QY	305	LLGGATPEVADAAYAVGRNLGLAQIVLDMLDYTSATDLGKPACADLOLGATAPALFA	364	:	:	:	:	:	:	:
Dd	185	ITSQSDKEFVRKLKMIGHYIGMSFOIIDVDLPFTSTEKKGAPVGSDDLNGHIITPILLE	244	:	:	:	:	:	:	:
QY	365	WGHNADELGMIR--KFSDPGDVERARELVEKSDGLEKTRALAEYAOKALAIARTFPES	422	:	:	:	:	:	:	:
Dd	245	MKNKDPFKLKLEQRRLDRSERKEFEECTQIIRKSIDSIDEAKAAVS KYLSALDPISELPPG	304	:	:	:	:	:	:	:
QY	423	PARKALEQULTDKVLTFRS	439	:	:	:	:	:	:	:
Dd	305	HPRKSLLSLTWKMGSRN	321	:	:	:	:	:	:	:

```

RESULT 6
US-09-815-242-5315
: Sequence 5315: Application US/09815242
: Patent No. US20020061563A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: fastSeq for Windows Version 4.0
: SEQ ID NO 5315
: LENGTH: 312
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-5315
Query Match      16.8%; Score 367.5; DB 10; Length 312
Best Local Similarity 26.5%; Pred. No. 1,4e-73;

```

Matches	97;	Conservative	68;	Mismatches	142;	Indels	59;	Gaps	4;
Qy	75	EMSNLTNSIRSLGSGHPSLDTVAKYVQSEGHIRPLMLVLLMAQATEVAPKVGMEKV	134						
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	4	EIKKVEQGLERAKIISKDSVLEQASLHLLSSGGKRRAPFVLLSSQ-----	48						
Qy	135	EVPNVNGLAPEVLNKNPDMNMRSGPLTYDGEIEGQTSINILASQRLAETEMHAAS	194						
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	49	-----FGKDEQSEQTYQV-----AAVALLEIHMAI	73						
Qy	195	LHHDVIDIASETTRNAPSGNQAFQKAKILAGDFILFGASVALARLNREVIETLLATYIA	254						
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	74	LHVDVIDIKSDRRKRGKLTISKKWDQTTLLTGNFLLGLGHEHMAVKONRVHQILSESIV	133						
Qy	255	NLVSEEPQQLKATYDDALAEATATQETFPYYIQKTYLTKASLIASCRASALLGGATPEVA	314						
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	134	DVCRGELFEQFODQF-----SQOTIINYLRIIRKRTALLIOLISTEVAITYSQSDKSTV	186						
Qy	315	DAAYAYGNLGLAQIYVDMLDYTSATDLGKRPAGADQLGLATPAPAFAMKHAHELSPM	374						
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	187	RKLKMIIGHYIMSPQIIDVDLFTSTEKKLRKPPVSSDILNHIIPLFILLEMRKNPDKFLK	246						
Qy	375	IK--RKSPDQGVGEARBLVEKSDLEKTRLAEEVAAKALDAITRPESPAPKALBOLT	432						
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	247	IEQLRDSERKEFEQCIIRKSDSIDAKAVSSKYLSKALDLISELDPGHPKSLLSLT	306						
Qy	433	DKVLTR 438							
		: : : : : : : : : : : : : : : : : : : : : : : : :							
Db	307	KKWGSR 312							

```

RESULT 7
US-09-815-242-10573
Sequence 10573, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zysekund, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITPA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PASTESEQ for Windows Version 4.0
SEQ ID NO 10573
LENGTH: 326
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10573
Query Match 15.8%; Score 347; DB 10; Length 326
Best Local Similarity 27.8%; Pred. No. 7.9e-22;

```

Matches 108; Conservative 56; Mismatches 146; Indels 78; Gaps 9;

QY 59 PDPNQLINLNVGPEMSNLTNSIRSLGSHPSLDTVAKYVYVQSGKHIRPLMLVILMA 118
 Db 9 PDVERSLTETCELJEOTLHTRNEDIEE-----ALVTLAR-----SGGKLLRPAPFFLFA 57

QY 119 QATEVAPKVGWEKVEVPVNEGLAPPEVLNDKNPMMNMRSGPLTKDGEIGOTSNILA 178
 Db 58 QL-----GDEEKEKQOLL 71

QY 179 SORRLAEITEMIHAASLLHDDVIDASETRRNAPSGNOAFGNKMAILAGDFLLGRASVALA 238
 Db 72 --KIASLEILHWATLIHDDIIDSPLRGAVTIQSGYKQVAVYTGDLJLFTFFFLIA 128

QY 239 RLRY-PEVIELLATVIANLVEGEFPMQKNVVDVDAIETATQETPDYLYQKTYLKTASLIA 297
 Db 129 DAMNGSEFMKINAQMKRLLLGELDOMSHRPFDRMSIPA-----YLRVNGKTAEFLS 181

QY 298 KSCRASALLGGATPEVADAAYVGRNIGLAFQIVDDMLDVTVSATDLGKPGADLQGLA 357
 Db 182 LSCLEGAYFGHSSKEVORLAKRIGRHIGIAPQVYDDILDYADTETLKKPALEDLSQGVY 241

QY 358 TAPALFAWKHAEGLPMIKRKFSPG---DVERARE---LVEKSDGLEKTRALAEVYAK 411
 Db 242 TLPLLFAY---QAAPDVFSYLDKGRAITLLEAEVAALVNDYHGVTEAQAFAKKVTNK 297

QY 412 ALDAIRTFPSPARKALEQLTDKVLTRS 439
 Db 298 AITDIOQLPDGTAKETLLSLTELLHRS 325

RESULT 8

US-09-738-626-4030
 ; Sequence 4030, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OKAZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patent in ver. 3.0
 ; SEQ ID NO 4030
 ; LENGTH: 350
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 ; US-09-738-626-4030

Query Match 15.5%; Score 340; DB 9; Length 350;
 Best Local Similarity 27.4%; Pred. No. 3.4e-21;
 Matches 115; Conservative 58; Mismatches 155; Indels 92; Gaps 14;

QY 35 RCTPT-SRPSSSMAAVSSASRLVEPPD-----NQPLINPLNVGPEMSNLTNSIRSL 87
 Db 5 RTVPTRSHGLGKEGVSTTGASQVEFGDPDLTARINDAMVQVEELLHTELS----- 54

QY 88 GSGHPSLDTVAKYVYVQSGKHIRPLMLVLLMAQAATEVAPKVGWEKVEVPVNEGLAPPEV 147

Db 55 -SGEDFLVDIVMLHTRAGGRFRPFMFALL---ASEFGEK-----PLSE----- 93

QY 148 LNDKNPMMNMRSGPLTKDGEIGOTSNILASQRRRLAEITEMIHAASLLHDDVIDASETR 207
 Db 94 -----NVI-----KAAVVEITHLATLYHDDVMDASMR 122

QY 208 RNAPSGNOAFGNKMAILAGDFLLGRASVALARLNPEVIELLATVIANLVEGEFPMQKN 267
 Db 123 RGVPSANARWDSNVAILAGDILLAHASGLMSQL-GTDTVAHFAETFGELVTG---QMRET 178

QY 268 VDDAIEATATQETPDYLYQKTYLKTASLIASCRASALLGGATPEVADAAYVGRNIGLA 327
 Db 179 V-----GPRDTPDPIETVNVIREKTGVLIASAGYLGAMHAGAPEHIDALKNFGAAGMI 233

QY 328 FOIVDDMLDVTVSATDLGKPGADLQGLATAPALFAWKH-----AELGPMIKRKFSPDG 383
 Db 234 FOIVDDIIDIFETHESGKTPGTDLREGVFTLPVLVALREDTPVGAELRDILTGPLEDDE 293

QY 384 DVERARELVEKSGLEKTRALAEY-----AQKALDAIRTFPSPARKALEQLTDKVLTR 438
 Db 294 TVNHVLELLSQSG--RQAALDEVYRYMDIANAELORL---PDSTVKEALRLNLTATFTVKR 348

RESULT 9
 US-09-925-637-66
 ; Sequence 66, Application US/09925637
 ; Patent No. US2002010338A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Polypeptides
 ; FILE REFERENCE: PB560
 ; CURRENT APPLICATION NUMBER: US/09/925,637
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/23773
 ; PRIOR FILING DATE: 2000-08-31
 ; PRIOR APPLICATION NUMBER: US 60/151,933
 ; PRIOR FILING DATE: 1999-09-01
 ; PRIOR APPLICATION NUMBER: US 08/781,986
 ; PRIOR FILING DATE: 1997-01-03
 ; PRIOR APPLICATION NUMBER: US 08/956,171
 ; PRIOR FILING DATE: 1997-10-20
 ; PRIOR APPLICATION NUMBER: US 60/009,861
 ; PRIOR FILING DATE: 1996-01-06
 ; NUMBER OF SEQ ID NOS: 74
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 66
 ; LENGTH: 273
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (261)..(261)
 ; OTHER INFORMATION: Xaa equals any amino acid
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (271)..(271)
 ; OTHER INFORMATION: Xaa equals any amino acid
 ; US-09-925-637-66

Query Match 15.5%; Score 339; DB 10; Length 273;
 Best Local Similarity 30.6%; Pred. No. 2.9e-21;
 Matches 85; Conservative 56; Mismatches 121; Indels 16; Gaps 3;

QY 160 SGPLTKDGEIGOTSNILASQRRRLAEITEMIHAASLLHDDVIDASETRRNAPSGNOAFGN 219
 Db 5 SSGFGKDEQTSQTYQV-----AVALDIHWATLHVHDDVIDKSKRRKGLTISKKWDQ 57

QY 220 KMAILAGDFLLGRASVALARLNPEVIELLATVIANLVEGEFPMQKNVTDVDAIEATATQ 279
 Db 58 TTAILTGNFLALGLEHLMAVKDNVRHQLISESIVDVCRGELFQFDQFN-----SQ 110

QY 280 TFDYLYQKTYLKTASLIASCRASALLGGATPEVADAAYVGRNIGLAFOIVDDMLDVT 339

Db 111 TIINYLRINKTALLQISTEVGAITSQSDKETVRKLMIGHYIGMSFOIIDVDFTS 170
QY 340 SATDLGKPGADLQLGATAPALFAMKHAELGPMIK--RKFSDPGVERARELVEKSDG 397
171 TEKLTGKPGVSDLNGHITLPILEMRKNPDKLIEQLRSDSRKFEFECEIQIRRSDS 230
QY 398 LEKTRALAEYAKALDAIRTFEPSPARKALEQLTDKY 435
Db 231 IDEKAVSSKYLKALNLISELPDGHPSRLXLSLTQKM 268
RESULT 10
US-09-815-242-11306
; Sequence 11306, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11306
; LENGTH: 307
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-815-242-11306
Query Match 13.8%; Score 302.5; DB 10; Length 307;
Best Local Similarity 29.4%; Pred. No. 4.2e-18;
Matches 97; Conservative 56; Mismatches 138; Indels 37; Gaps 11;
QY 119 QATEVAPKVGMEKVEVPPVNEG---APPEVINDKNPDMNMNRSGPLTKDGEIEGQTS 174
Db 5 QKTIQKIKASWIKIEISAFIDELFSKIGPKMKLSKL--MLAL-----LNKXTD 52
QY 175 NILASQ--RLAETEMTHASLLHDDVYIDASETRRNAPSGNQAFGNKMAILLADFLIGRA 233
Db 53 AILLDKAFNLCALIVEMIQTSLLHDVDYIDKAMMRKLPSTINALFGNNAWMLGDFVFSKA 112
QY 234 SVALARLNPEVIELLATVIANLVGEFEMQKNTVDDAIEATATQETFDYVLOKTYLTKTA 293
Db 113 PFELISK--GSIASLSNAVRLSRGE-----IEDVPVGGSGNSKQKTYMRILEKTA 164
QY 294 SLIAKSGRASALLGATPEVADAAYA--YGRNLGLAFQIVDDMLDYVSATDLGKPGADL 352
Db 165 HFTFASLKSMAILLNKDAKM---YADFGNLFQWAFQIIDDLDITQDAKTLGKPNSSDF 220
QY 353 QLGATAPALFAMK--HHAELGPMIKRKFSDPGD--VERARELVEKSGLEKTRALAEYA 409

Db 221 KECKTTLPIILLVEKLNQDGLLSYFKQDSHEIEMTEKFKQYGTIEETLKIAQVYS 280
QY 410 QKALDAIRTFEPSPARKALEQLTDKYLTRS 439
Db 281 KKALEAIK-----GENNLLIEKLAQDVIVRT 306
RESULT 11
US-10-108-915-26
; Sequence 26, Application US/10108915
; Patent No. US20020177204A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Shen, Jennie
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: Geranylgeranyl Pyrophosphate Synthases
; FILE REFERENCE: BB1286 US NA
; CURRENT APPLICATION NUMBER: US/10/108,915
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/452,238
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/110,592
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 26
; LENGTH: 367
; TYPE: PRF
; ORGANISM: Glycine max
US-10-108-915-26
Query Match 12.4%; Score 272; DB 9; Length 367;
Best Local Similarity 23.8%; Pred. No. 2e-15;
Matches 106; Conservative 59; Mismatches 152; Indels 128; Gaps 12;
QY 19 IASVTLRTASAPSL-----RLRCTPTR-----PSSGMA-----AAVSS 52
Db 1 MSAVNLNTWPRPSFTLNGQATRRSSPSHFHGVNKLPSPISSLTVAKRSFTLSAVLT 60
QY 53 ASRLVEPDPNQLPLNLNVGPEMNLTSNIRSLGSGHP--SLDTVAKYVYQSEKHIRP 111
Db 61 KEDYVETBEKPIIPDKYMYVSKASAVNKALDDAVSLREPKIHAMKXSLAGKRYRP 120
QY 112 LNVLLMAQATEVAPKVGMEKVEVPPVNEGALPPEVLNDKNPDMNMNRSGPLTKDGEIEG 171
Db 121 VLCV-----AACELVGEEATAMPA----- 140
QY 172 QTSNIIASQRLAETEMTHASLLHDDV--IDASETRRNAPSGNQAFGNKMAILLADFL 229
Db 141 -----ACAETEMTHWSLLHDDLPCMDNDLDRGKPTNKKVGEDEVAVLAGDAL 188
QY 230 LGRA---SVALARLNPEVIELLATVIANLVGEFEMQKNTVDDAIEATATQETFDYVLO 286
Db 189 LAFAPENHIASTRGASPRIVATIGELARSTGSEBLVAGQVVDINSEGLA-----DVULE 243
QY 287 KTYL-----KTASLIASGRASALLGATPEVADAAYAAYGRNLGLAFQIVDDMLDYVSAT 342
Db 244 RLEFIHVHKTALLGCAVVALIGGTFDDEVEKLRKFRARYIGLLFYVDDILDVTKSSQ 303
QY 343 DLGKPGADLQLGATAPALFAMKHAELGPMIKRKFSDPGDEVERARELVEKSGLEKTR 402
Db 304 ELGKTAGKDLVADKVTYPLT-----GIEKSK 330
QY 403 ALAEYAKALDAIRTFEPSPARKA 427
Db 331 VPAKLNKDAQDQVGF--DPYKAA 353
RESULT 12
US-09-815-242-10069
; Sequence 10069, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10069
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10069

```

```

Query Match 11.9%; Score 260; DB 10; Length 299;
Best Local Similarity 27.8%; Pred. No. 1.6e-14;
Matches 114; Conservative 45; Mismatches 104; Indels 152; Gaps 19;

QY 49 AVSSASRLVDPDP--NQPLINPLNLVGPMSNLTNIRSLGSGHP-SLD 106
DB 14 ANQALSRIAPLPQNTPVVE-----TMOYGALLG-----G 44

QY 107 KHIRPLMVLMAQATEVAPKVGQWKEVVPVNEGLAPPEVLNDKNPMMNMRSGPLTKD 166
DB 45 KRLRPFLV---YAT----- 55

QY 167 GEIGQTSNILASQRRRLAEITEMIHAASLLHDDV--IDASETRNAPSGNQAFGNKMAIL 224
DB 56 GHMFGVSTNTLDAP---AAAVECIHAYSLIHDDLPAMDWDDLRRLGLPTCHVKFGEANAIL 112

QY 225 AGDFLLGRASVALARLNPEV-----IELL-----ATVIANLVEGEFMQL-----KNTVDD 270
DB 113 AGDALQTLASILSDADWPEVSDRISMSISELASAGIACMGCGQALDLDAEGKHVPLD 172

QY 271 AIEATATQETFDYLYQKTYLKTASLIKSCRASALLG-----ATPEVADAAYAGRNGLG 325
DB 173 ALERIHRRH-----KTGALIRAAVRLGALSAGDKGRALP-VLD---KYAESIG 216

QY 326 LAFOIVDDMLDYTVSATDLGKPGADILQGLATAPALFAWKHAEELGPMIKRKFSDPGDV 385
DB 217 LAFOVQDDILDVTVSGTATLGRQADQDLGKSTYPAL-----LGLEQARK----- 261

QY 386 ERARELVKSDGLEKTRALAEVYAKQALDAIRTPESPARKALEFQTDKVLTRSR 440
DB 262 -KARDLID--DARQLKQLAEQ-----SLDT-----SALEALADYIIQRNK 299

RESULT 13
US-10-108-915-22
; Sequence 22, Application US/10108915
; Patent No. US20020177204A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Shen, Jennie
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: Geranylgeranyl Pyrophosphate Synthases
; FILE REFERENCE: BB1286 US NA
; CURRENT APPLICATION NUMBER: US/10/108,915
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/452,238
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/110,592
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Glycine max
; US-10-108-915-22

Query Match 11.6%; Score 255.5; DB 9; Length 350;
Best Local Similarity 23.4%; Pred. No. 4.7e-14;
Matches 94; Conservative 59; Mismatches 141; Indels 107; Gaps 11;

QY 37 TPTSREPSSWAAVSSASRLVDPDPNQPLINPLNLVGPMSNLTNIRSLGSGHP-SLD 95
DB 33 TVTKRRAFSLSAVLT-----VETEERPPIDFKNYMLSKASAVNKGLDDSVLSUREPKKH 87

QY 96 TVAKYVQSEGKHIRPLMVLMAQATEVAPKVGQWKEVVPVNEGLAPPEVLNDKNPDM 155
DB 88 EAMRYSLLAGGKVRPVLV-----ACELVGGHEATAMPA----- 123

QY 156 MNMRSGPLTKDGEIEGOTSNILASQRRRLAEITEMIHAASLLHDDV--IDASETRNAPSG 213
DB 124 -----ACALEMIHTMSLIHDDLPCMDNDLRRGKPTN 155

QY 214 NQAFGNKMAILAGDFLLGRASVAL--ARLNPEVIELLATV--IANLVEGEFMQLKNTVDD 270
DB 156 HTVFGEDVAVLAGDALLAFEHIAASTRGASAPRILRAITGELARSIGSEGLVAGQVVDI 215

QY 271 AIEATATQETFDYLYQKTYL-----KTASLIKSCRASALLGSGATPEVADAAYAGRNGLG 326
DB 216 NSBGLA-----DVGLERLFEFHVHVKTAALLEGAVVLGAILGGTDDDEVEKLRKFAYIGL 270

QY 327 AFOIVDDMLDYTVSATDLGKPGADILQGLATAPALFAWKHAEELGPMIKRKFSDPGDVE 386
DB 271 LFOVQDDILDVTVSGTATLGRQADQDLGKSTYPKLL----- 307

QY 387 RARELVKSDGLEKTRALAEVYAKQALDAIRTPESPARKA 427
DB 308 -----GIBKSKEFAAKLNKDAQDQLAGF--DFVKAA 336

RESULT 14
US-10-108-915-14
; Sequence 14, Application US/10108915
; Patent No. US20020177204A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Shen, Jennie
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: Geranylgeranyl Pyrophosphate Synthases
; FILE REFERENCE: BB1286 US NA
; CURRENT APPLICATION NUMBER: US/10/108,915
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/452,238
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/110,592
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 232
; TYPE: PRT

```

ORGANISM: Oryza sativa
US-10-108-915-14

Query Match 11.6%; Score 254.5; DB 9; Length 232;

Best Local Similarity 29.3%; Pred. No. 3.1e-14;

Matches 75; Conservative 39; Mismatches 93; Indels 49; Gaps 6;

QY 187 TEMHAASLHDDV--IDASTRRNAPSNGOAFGNKAAIILAGDFLGRASVALARL----- 240
DB 1 TSIQHTMSLVHDDLPCHDDDDLRGKPTCHVVGEPFIAVLTGALLSLSPHMAARFDSYP 60

QY 241 -----RNPEVIELLATVINLVGEFMOQKNTVDDAIEATATQETFDY-YLOKTYL-KT 232
DB 61 PDIDADKPARVRAIETELARCISEGIVAGQVVD--LEMTGSTETVPLERLBIYHLHKT 118

QY 293 ASLIAKSCRASALIGATPEVADAAYAGRNGLAFOIVDDMLDYTSATDLGKPAQADL 352
DB 119 AALLEAVVIGAILGGSDQIESLRMYARSIGLLFQVVDILDVTKSSELGKTAGKDL 178

QY 353 QLGATAPALFAMKHAELGPMIKRKSPPGDVERARELVEKSDGLEKTRALAEIYAOKA 412
DB 179 ASDKTYTPKLL-----GLEKSRFEAKLISDA 205

QY 413 LDAIRTFESPARKAL 428
DB 206 REQLSGFDQETAPLL 221

RESULT 15
US-10-108-915-18

Sequence 18; Application US/10108915
Patent No. US20020177204A1

GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Shen, Jennie
APPLICANT: Williams, Mark

TITLE OF INVENTION: Geranylgeranyl Pyrophosphate Synthases
FILE REFERENCE: B01286 US NA
CURRENT APPLICATION NUMBER: US/10/108,915

PRIOR FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/452,238
PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-01

PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/110,592
PRIOR FILING DATE: EARLIER FILING DATE: 1998-December-02
NUMBER OF SEQ ID NOS: 46

SOFTWARE: Microsoft Office 97
SEQ ID NO 18
LENGTH: 369
TYPE: PRT
ORGANISM: Glycine max
US-10-108-915-18

Query Match 11.5%; Score 253; DB 9; Length 369;
Best Local Similarity 25.2%; Pred. No. 8.3e-14;

Matches 105; Conservative 54; Mismatches 163; Indels 94; Gaps 11;

QY 59 PDPNOPLINPLNVGPEMSNLTSNIRSLGSHPSL-----DTVAKYVQS 104
DB 10 PRPTSMNTSHTLPPHPTLTTLTKLASGTPKLISSPFLVSAPVTKHVTYTGQIQ- 68

QY 105 EGKHIRPLMLLMAQATEVAKVQGEKVEVPVNEGLAPPEVLNDKPNPMNRRSGPLT 164
DB 69 -----LQDPPLNDFPKYMIAKAHTVNOALDAIALRDPKHIOAMRYSLLA 115

QY 165 KDG-----EIGQTSNIIASQRLAETEMTHAASLHDDV--IDASTRRNAP 211
DB 116 GGRKVRPVLICIAACELVGTGTE--ATAIPACAVEMTHMSLIHDDLPCHDNDLRRGKP 172

QY 212 SGNQAFGNKAAIILAGDFLGR-----SVALARLNPEVIELLATVIA--NLVEGEFM 262
DB 173 TNHKVYGEDVAVLAGDALLAFAPENVAASTEGVSPSRVRAIGELAKSIGTBGLVAGQVV 232
QY 263 QLNKTVDDAIEATATQETFDYIYLOKTYLKTASLIAKSCRASALIGATPEVADAAYAGR 322

DB 233 D-----IDSEGVANVGLTLEFI---HVKTAALLAAVVLGAIVGGSDDEVEKLRKRFAR 285
QY 323 NGLAFQIVDDMLDYTSATDLGKPAQADLQGLATAPALFAMKHAELGPMIKRKSFPD 382
DB 286 CIGLFQVVDILDVTKSSELGKTAGKDLVADKTYPKLL----- 326

QY 383 GDERARELVEKSDGLEKTRALAEIYAOKALDAITTFESPARKALEQITDKVLTTR 438
DB 327 -----GIDKSKEFAQELLDAKEQLSGF-DPPKAAPLFALTNYIAYR 367

Search completed: January 16, 2003, 09:56:38
Job time : 21 secs


```

Db      4  RSGAAHHIKLISRRRCRFKSPFAVLNMAASKLVT-----KILMNNPISIVSKEMTTLAKN  59
Oy      83  IRSLLSGHSPSLDYTAKEYVQSEGHIRPLMLVLLMAQATEVAPKQ-----GHKRYEV  136
Db      60  IVALIGSHPIVNLKTSYFETEGKKVREPLVLLLSRLSALSEIPMERHNLKIDKSDVED  119
Oy     137  PV-----NEGLAPEPELVNDKVPDMMNRSGLPTKDGEEIEGOTSNIILASQRL  183
Db     120  PLYSKPSONOLFQRPASSISPLHILHGIR- LNP LTKGPEPLPEETFPKQRIPLKQRL  178
Oy     184  AEITEMIHAASSLHDDVIDASETRRNAPSGNAFQNKWAILAGDFPLGRASVALARLNP  243
Db     179  AEIVEMIHHTASLLHDDVIDHSDTRGRPGSGNAAFNNKAVLAGDFELGRATVISRLHP  238
Oy     244  EYIELLATVIALVGEFQNLKQTVDDAEAT-----  275
Db     239  EYVELMSSINIALVGEFQNLKNTSIDADITIEHGHKLLPVSKKLEVEKEHDFVPSRQ  298
Oy     276  -----ATOETPEYVUKTYTLPASLIASKCRASALLGGATEPEVADAAVAYGRNLG  326
Db     299  OGLQLSHDOIETAEPEYIHKTYLKTALISKSCCAALISASPAVVIDECDFPRNLGI  358
Oy     357  AFQIVDDMLDYTVSATDLGKPRAGADLQUGLATAPALFAWKHHAEIGPMIKKXSPDGVE  386
Db     359  CFQVLDDMLDFTVSGCDLGKPSGADLKIGIATAVPLFAKEDPDSGLPISRNFSGDVE  418
Oy     387  RARELVEKSDGELKTRALAEYAKALDAIR--TFPESPRARKLEQDTDKVLTRSR  440
Db     419  KTIIDVRLHNGIAKTKILAEYRDALQVLRJSDSEDSRSLLEPLVLSILTRRK  473

```

RESULT 2
JC5429
di-trans,poly-ists-decyprenylcitransferase (EC 2.5.1.31) [similarity] - fission yeast (S)
N/Alternate names: decaprenyl diphosphate synthase
C/Species: Schizosaccharomyces pombe
C/Date: 17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 28-Jul-2000
A/Accession: JC5429; T50407
J/Stucki, K.; Okada, K.; Kamiya, Y.; Zhu, X.F.; Nakagawa, T.; Kawamukai, M.; Matsumura, H.
J. Biochem. 121, 496-505, 1997
A>Title: Analysis of the decyprenyl diphosphate synthase (dps) gene in fission yeast sug
A/Reference number: JC5429; PMID:97279041; PMID:9133618
A/Accession: JC5429
A/Molecule type: DNA
A/Residues: 1-378 <SU>
A/Cross-references: DBDJB:D84311; NID:G1845554; PID:BAA12314.1; PID:G1845555
R/McDonough, R.C.; Rajendram, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.
submitted to the EMBL Data Library, January 2000
A/Reference number: Z25059
A/Accession: T50407
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-378 <MCD>
A/Cross-references: EMBL:AL136498; PIDB:CAB66154.1; GSPDB:GN00067; SPDB:SPBPJ694_01
A/Experimental source: strain 972h(-)
C/Comment: This enzyme plays a role in the biosynthesis of ubiquinone which functions as
the 10-isoprene-unit side chain of ubiquinone.
C/Genetics:
A/Gene: dps; SPDB:SPBPJ69_01; SPDB:SPBPJ694_01
A/Map position: 2L
A/Introns: 325/3
C/Superfamily: prenyl transferase A
C/Keywords: transferase

	Query Match	Best Local Similarity	Matches 191;	Conservative 60;	Mismatches 96;	Indels 27;	Gaps 3;
Oy	67	NPLNLVGPMSU...LTSLSGSHPSLDPTVAKYVGOSGEKRIIRPLEMLYLMAQATEVPX 126	41.5%; Score 910.5; DB 2; Length 378;				
Db	32	NASHIKNELEISFOICKOMLNSENSEFLECKSKYTTIADGGKMRSTVLNSKAISLSCHG 91	51.1%; Pred. No. 5.1e-56;				
Oy	127	VQGWEKVEVPEINEGLAPEEVLNDKNPDMMNRSCPLTYKGIEIGQTNNILASQRRLAEI 186					

```

Db      92 ID-----RSVSGEKYIDDDDLKS-----FSTGQLPESQRLAQI 125
QY      187 TEMIHAASLLHDDVIIDASETRRRNAPSGNOAFGNKMAILACDFLLGRASVALARLRNEVIT 246
      126 TEMIHISALLHDDVIDIHANVRKRGSPSSNVAFGNRRSIIAGNFILIRASTAMALRNPQYT 185
QY      247 ELIATVIANIVLVEGRFMOLKUNVDIAIATATNOETFDVYLKQTYLKTASLAKSCRASALL 306
Db      186 ELIATVIANIVLVEGRFMOLKUNVDIAIATATNOETFDVYLKQTYLKTASLAKSCRASALL 244
QY      307 GGAATPEVADAAYAAVAGRNILGIAFOIIVDDMLDYVSATDUGKPGAGADLOLGATAPALPAMK 366
Db      245 GCGSPYATATANGVEYGRGICIGTAFQJLMDVDVLDYTSKDDDTLKGAAAGDLKLGATAPVLPAMK 304
QY      367 HHAEILGPMIKRPSDPCGVDERARELVEKSDGLEKTRALAEVYOKALDAIRTPESPAPAK 426
Db      305 KYPEIGAMIVNRFMHPESDIOGRASIVLECTDAIBQITITWAKEVIKKAKDSLCLCPDSPARK 364
QY      427 ALLEQTLDTKVLTRSR 440
Db      365 ALPALADRVITRRK 378

```

```

RESULT 3
T43193
trans-pentaprenyltransferrase homolog - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 11-Jan-2000 #sequence_version 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43193
R:Yoshida, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA:Res. 4, 369-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: Z17323; MUID:96162722; PMID:9501991
A:Accession: T43193
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-356 <YOS>
A:Cross-references: EMBL:D89265; NID:91749737; PIDN:BAAL1392.1; PID:G1749738
A:Experimental source: strain PR745
C:Superfamily: prenyl transferase A

Query Match          37.5%; Score 823.5; DB 2; Length 356;
Best Local Similarity 50.0%; Pred. No. 5.5e-50;
Matches 175; Conservative 56; Mismatches 92; Indels 27; Gaps 4;

Oy 67 NPLNVGEMENLTNINISLGSCHPSLDVTAKYVJSEGGHIRPLMWLMAQATEVAPK 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 32 NASHLRKHELQISGICIGOMLNSNSEFLECSKYTTIAGQMRSLVLLMSKATSL--- 88

Oy 127 VQGWKEVVEPVNBSGLAPEVLYNDKNPDMNMNRSGPLTKDGEIEGQTSNIIASQRRLAET 186
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 FPGIDR-----SVGDKYTDDDLRS-----FSTQIIPSQLRLAQI 125

Oy 187 TEMTIAASLHDDVDVDASETRRNAPSGNQAQGNKALIIAGPFLGRASVALARLNPETV 246
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 TEMHIAALLHDDVDVDAHNVRRSGSSSNVAAQGNRSIIAGNFIARASTVWARLENPQVT 185

Oy 247 ELLATVIVNLVEGGEFMQKNTVDDAIEATATQEFDFYLOQTYLTKTASTLAKSCRASALL 306
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 ELLATVIVNDLVRGELTQAKNTMDPS-SLEIRQSNFDYIIRISFLKTAISLISKSCASTILL 244

Oy 307 GGATPEVADAAYAYAGRNIGLAFQIVDDMLDYTSATDLGKPDAGDLOGLATAPALPFAMK 366
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 GGCSPTVATAAGEYGRICGTAFQGLMDVDLVDTYSKDDTLGKAAGADLKGLATAPALPFAMK 304

Oy 367 HHAELGPIIKKKFSDPGVEPARBELVEKSDGLEKTRALAEYAQALDAI 416
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 KYPELGAMIVNRFNHPISDIOQARSIVECSDAIEOPITWAIYIKKRPSPV 354

```

geranylgeranyl pyrophosphate synthase crtE homolog - Synchocystis sp. (strain PCC 6803)
N;Alternate names: Hypothetical protein slr0611
C;Species: Synchocystis sp.
A;Variety: PCC 6803
C;Date: 14-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
R;Accession: S76966; S74427
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis
sp.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76966
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-67 <KAN>
A;Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAA20922.1; PID:g342616
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
A;Accession: S74427
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 68-323 <KAW>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAA16579.1; PID:g165165
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: crtE
C;Superfamily: prenyl transferase A

Query Match 30.8%; Score 676.5; DB 2; Length 323;
Best Local Similarity 40.8%; Pred. No. 8.7e-40;
Matches 153; Conservative 63; Mismatches 102; Indels 59; Gaps 4;

QY 65 LINPLNLVGP---EMSNLTNSIRSLGSGHPSLDTVAKYYVQSEGKHIRPLMWLLMAQAT 121
Db 1 MISTTSLFAPVDQDLRLTDNLKELVGARHPILGAARAEHLFEAGCKVRPAIVLLVSRA 60
QY 122 EVAPKVGWEKVEVPVNEGLAPEVLNDKNPMMNMRSGLTKDGEIEGQTNSILASQR 181
Db 61 -----LLDQ-----ELTAHR 71
QY 182 RLAEITMIHAASLLHDDVIDASTRRNAPSNGNOAFGNKMAILAGDFLGASVALARLR 241
Db 72 RLAEITMIHTASLVHDVDVEDADLRNVPTVNSLFONRVAVLAGDFLFAQSWWYLANLD 131
QY 242 NPVEIILLATVIANLVGEFGMLKNTVDDATEATQCTFDYLYOKTYLKTSIAKSQR 301
Db 132 NLEVVKLSSEVIRDFAGEILQSNRF-----TTDTLEYLEKSYFKTASLIANSK 184
QY 302 ASALLGGATPEVDAAYAGRNLGAFQIVDDMDLYTVSATDLGKPAGADQLGLATAPA 361
Db 185 AAGVSDAPRDVCDHLVEYGKHLGLAQFIVDDILDFTSPTVELGKPGAGSDLISGNITAPA 244
QY 362 LFAWKHHAELGPMIKRKFSDDGVVERARELVKESDGLEKTRALAEYAQAQKALDIRTFPE 421
Db 245 LFAMEKYPLLKGLIEREFAQGDLEQLAELEVEQGDIIRSRRELAANOQAQLARQHLSVLEM 304
QY 422 SPARKALEQLTDKVLTR 438
Db 305 SAPRESLLELVYVIGR 321

RESULT 5
A40433
prephtyoene pyrophosphatase dehydrogenase (crtE) homolog - Cyanophora paradoxa
C;Species: Cyanophora paradoxa
C;Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 24-Sep-1999
R;Accession: A40433
R;Michalowski, C.B.; Loeffelhardt, W.; Bohnert, H.J.
J. Biol. Chem. 266, 11866-11870, 1991
A;Title: An ORF323 with homology to crtE, specifying prephtyoene pyrophosphate dehydroge
A;Reference number: A40433; MUID:91268060; PMID:1711042
A;Accession: A40433
A;Status: preliminary

[illegible]

Db	77	TEIHTASTASVHDDIDDESIVRKGIPTVHSDPFGTKAIILAGDPLFAGSSMYLANLESLEV	136
Qy	247	ELLATVIANLVGEFMQLKNTVDDAI EATATQETFDYIIOKTYLKTASLIASCASALL	306
Db	137	KLSKVIDTDFAGEGIR-----GNOFKVDLTLEEYLEKSFYKTASLIASSFAALL	189
Qy	397	GGATEVADAAAYARNLGLAFQIVDMMDYTVASATDLCKPAGADLOGLATAPALFAMK	366
Db	190	SHVDLTVANADLTNYGSHLGLAFQIVDDI LDFSTSTEELOKPCSCDPLKKNLTAPALFALE	249
Qy	367	HHAELGPMTKRKSPGDVERARBELVEKSDGLEKTRALAAEYAOALDIRFPESPAPK	426
Db	250	ONSELPILQRFSPSEKDEFTYLOIVEETKALEKRELLAMEHAOVAIOCLEMLPPSSSKE	309
Qy	427	ALBQLTDKVLTR 438	
Db	310	ALKLITKVLER 321	

RESULT 7
 AH1818
 solanesyl diphosphate synthase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AH1818
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriquchi,
 Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH1818
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-323 <KUB>
 A:Cross-references: GB:BA000019; PIDN:BA877620.1; PID:gl7135074; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: sds
 C:Superfamily: prenyl transferase A

```

Query Match          30.2%  Score 661.5; DB 2; Length 323;
Best Local Similarity 38.4%  Pred. No. 9.7e-39;
Matches 145; Conservative 72; Mismatches 102; Indels 59; Gaps 4;

Dc      61 PNOPIINPLNLGPEPMSNLTNSIRSLGSGHPSLDVTAKYVYQOSEGKHIRPLMVLMLMQA 120
3 PATSLFTF---VEADLRILADNLKQVLGNRHPILPAAEAHLFGAGGKKRIRPAIVLLISRA 59

Qy      121 TEVAKRGVGMKEKVEVPVNNEGLAPEEVLNDKKPDMNNRSGLTGDEIGQTSIIIASQ 180
60 T-----MDQGITP-----RH 70

Qy      181 RLRLAETEMIIHAASLIHDDVIDASETRRNAPSGNOAFGNKAILAGDFLLGRASVALARL 240
71 RLRLAETEMIIHAASLIHDDVIDASETRRNAPSGNOAFGNKAILAGDFLLGRASVALARL 130

Qy      241 RNPEVIELLATVIANLVESEFMQLKNYVDALAEATATQETFDYLYQKTYLKTASIIAKSC 300
131 DNLQVKKLSEIYMDLATEIGQGNRFPASI-----SIFTEYIEKSYVKTASIIANSS 183

Qy      301 RNSALLGCTPEVADAAYVGRNLGLAFQIVDDMLDYTSATDLKPKPGADQLQLATAP 360
184 KQAGLISEVSPTEAHLVYGRNLGLAFQIVDDMLDYTSATDLKPKPGADQLQLATAP 243

Qy      361 ALFAMKHAALGPMIKRKTSDPGDERARELVEKSDGLEKTRALAEYAQKALDAIRTFP 420
244 VLFALAEKRYLEVLIEREPAGQDLEQALELIDQDSGIIQSSRELAHHTKLAIEHLATLP 303

Qy      421 ESPARKALEQLTDKVLTR 438
304 PSEHQALIKIAEVAISR 321

```

RESULT 8
S73189
prenyl transferase A - red alga (Porphyra purpurea) chloroplast
C:Species: Chloroplast Porphyra purpurea
C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 24-Sep-1999
C:Accession: S73189
R:Reich, M.; Munholland, J.
P:Plant MOL. Biol. Rep. 13, 333-335, 1995
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.
A:Reference number: S73108
A:Accession: S73189
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-923 <REI>
A:Cross-references: EMBL:U08804; NID:G127665; PIDN:AC08154.1; PID:G1276734
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C:Genetics:
A:Gene: preA
A:Genome: chloroplast
A:Superfamily: prenyl transferase A
C:Keywords: chloroplast

```

Query Match      27.6%; Score 604.5; DB 2; Length 323;
Best Local Similarity 36.1%; Pred. No.9,3e-35;
Matches 135; Conservative 69; Mismatches 111; Indels 59; Gaps 3;

QY    LINLINLVGPEMSNLTNIRSLSGSPSLDTVAKYVVOSEGRHIRPLMWILMAQAATEVA 124
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     7 LPHPIE---KELYSVEHNLSKVAGTRHPIIYAANKHLFEAGCGRLRPAILVAKST--- 60

QY    125 PKVGWKEVVEVPVNEGFLAPDEVLNDKNPDMNMNRSGPLTKDGEIEGQTSNLIASQRILA 184
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     61 -----SEQUOEIKPGORILA 74

QY    185 EITEMIHAASSLHHDDVDADSETRRNAPSNGOAFGNKKALLIAGDFLLGRASVALARLNPE 244
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     75 EITEIHIFHASLVHDVDIDECSTRGEXTKHKLFTNTKAVIALVAGLPFAQQSSWYLANIGNE 134

QY    245 VIELLATVIANLVGEEFMOLKNIVDDAIEFATATOETFPEDYLQCTLYKTASTLIAKSCRASA 304
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     135 VVKVITKVITDPFAGGELRQGLVHFDPSE-----SIDDVIEKSIFYTASHLVVASCRGA 187

QY    305 LLGGATBEVADAAYAYGRNIGLAFOIVDDMLDYTVSATDLGPKAGADLQLGLATAPALEFA 364
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     188 MLNDLNSQMNHDDLTYKGHMGLAFQIWDVDVLDIAGSKTSLGKPSGADPMNGNLTAPILFA 247

QY    365 WKHHAELGPMTKRKFSPPGVVERARELVEKSDGLEKTRLALAEYAQCALDAITFPESPRA 424
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     248 LTQGKLDQLQREFSPSERDISLALFLIKKSGGITAKDKAQVQAALCLOFLPESAP 307

QY    425 RKALEQLTDKVLTR 438
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     308 VSSLKETLTHPTLR 321


RESULT 9
G87775
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C/Acession: G87775
R:anonymouse, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; WUID:19069613; PMID:9651916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A>Note: published extrata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G87775
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <STO>
```

DB 132 ELITTEHRAELAEILEMIIHTASLIHDDVLESDMRRGKETVHELFGTRVAVLAGDFNFAQAS 191

Db 310 LREIISEFCFPGSLEEAIEIVNRGGTIKKAQELAKEKAELAKNLNCLPRSGFRSALED 369

Qy 431 L 431

Db 370 W 370

RESULT 12

AB84759

probable trans-prenyltransferase (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 24-May-2001

C/Accession: AB84759

R/Lin: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.

Neus, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: AB84420; MUID:20083487; PMID:10617197

A/Accession: AB84759

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-297 <STO>

A/Cross-references: GB:AE002093; NID:G3128225; PIDN:AA26705.1; GSPDB:GN00139

C/Genetics:

A/Gene: At2g34630

A/Map position: 2

C/Superfamily: prenyl transferase A

Query Match 25.6%; Score 561; DB 2; Length 297;
Best Local Similarity 37.5%; Pred. No. 8.9e-32;
Matches 134; Conservative 61; Mismatches 88; Indels 74; Gaps 6;

QY 92 PSLDTAKYVQS--EGKHRLPLVLLMAQATEVAPKVGMEKVEVPVNEGGLAPPEVLN 149

Db 7 PKASAAEYFEKRGVQKQFRSTILLMATALNV-----RVEPALI 47

QY 150 DKXPDMMNMSGPLTKXGIEGQTSNLAQRRLAETEMIHASLHDVDIVASERRN 209

Db 48 GSESDIV-----TSELVRORGAETEMIHASLHDVDIVADRRG 91

QY 210 APSGNQAFGNKMLADPFLGRASVALARLNPEVIELLATVIANLEGEFQNLKTV 269

Db 92 VGSINVMGK-----VVALATVENVHVTGETMEITS--- 124

QY 270 DAIEATATQETFDYLYLQTKYLTASLIKSCRSALIGATPEYADAAYVGNLGLAFQ 329

Db 125 ----STEQRYSMDYMQKTYKTSLSNSCKVAVALTGQTAEVAVALAFEGNLLGLAFQ 180

QY 330 IVDMDLYTASATDLGKPGADLQGLATPALFAWKHAELGPMIKRSPDGDVERAR 389

Db 181 LIDILDFGTISLGSJSDIRHGVTAPILFAWEFFQLEKVVQVEKDPKRVNDIAL 240

QY 390 ELVEKSGLEKTRALAEVAKALDAIRTPESP-----ARKALEQTLTKVLRSR 440

Db 241 EYLKSGKIGIRARLAMEHNLAAALIGSLPTDNEDVKKSRRLIDLTRVITRKK 297

RESULT 13

E65109

occiprenyl-diphosphate synthase (EC 2.5.1.-) - Escherichia coli (strain K-12)

C/Species: Escherichia coli

C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #ext_change 01-Mar-2002

C/Accession: E65109; PV0010

R/Lin: F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: E65109

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-323 <BLAT>

A/Cross-references: GB:AE000399; GB:U00096; NID:G2367201; PIDN:AA26219.1; PID:G1789578;

A/Experimental source: strain K-12, substrain W61655

R/Choi, Y.L.; Nishida, T.; Kawamukai, M.; Utsuni, R.; Sakai, H.; Komano, T.

J. Bacteriol. 171, 5222-5225, 1989

A/Title: Cloning and sequencing of an Escherichia coli gene, nlp, highly homologous to l

A/Reference number: J00027; MUID:89359178; PMID:2670911

A/Accession: PV0010

A/Molecule type: DNA

A/Residues: 11,156-157,196-323 <CHO>

A/Cross-references: GB:X68873; NID:G42128; PIDN:CAA4873.1; PID:G388220

C/Genetics:

A/Gene: nlpB

A/Map position: 69 min

C/Superfamily: prenyl transferase A

C/Keywords: transferase

Query Match 20.6%; Score 451.5; DB 2; Length 323;
Best Local Similarity 31.0%; Pred. No. 4.4e-24;
Matches 116; Conservative 52; Mismatches 143; Indels 63; Gaps 4;

QY 71 LVGPMSNLTNSIRSLSGHPSLDTVAKYVQSEGKHRLPLVLLMAQATEVAPKVGW 130

Db 9 LTAQDMAGVNAALIEQLNSDVQLINQLGYTVSGGKIRIMVLAARAV-----GY 61

QY 131 EKVEVPVNEGGLAPPEVLNDKXPDMMNMSGPLTKXGIEGQTSNLAQRRLAETEMI 190

Db 62 EGNAAVTI-----AALIEFI 76

QY 191 HAASLHDVIDASETRRNAPSGNQAFGNKMLADPFLGRASVALARLNPEVIELLA 250

Db 77 HTATLLHDVDVDESDMRGKATANAAGNAAVLDGPIYTRAFOMTSLGSLKYLEVMS 136

QY 251 TVIANLVGEFQNLKTVDDAIEATQETFDYLYLQTKYLTASLIKSCRSALIGAT 310

Db 137 EAVNVIAEGEVLQNMNVDPDI-----TEENYRWYVSTALFEAAACSGSLAGCT 189

QY 311 PEVADAAYAGNGLAQIIVDMLDYTSATDLGKPGADLQGLATPALFAWKHAEL 370

Db 190 PEEKGLDYGKYLGTAFQILDLDLYNADGEQLKXVGDNDLNGSKPTPLHMHHTP 249

QY 371 LGPMIKRSPDGD-----VERARELVEKSGLEKTRALAEVAKALDAIRTPESPARK 426

Db 250 EQAQMIRTAIEQGNRHLLFVLKEMNACGSLWTRQAEAEADKAIALQVLDPTWRE 309

QY 427 ALBQTLTKVLRSR 440

Db 310 ALIGLAHIAVQDR 323

RESULT 14

B91137

occiprenyl diphosphate synthase (imported) - Escherichia coli (strain O157:H7, substrain

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 03-Aug-2001

C/Accession: B91137

R/Lin: T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gatanara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: B91137

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-323 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA837489.1; PID:G13363539; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RMD 0509952

C/Genetics:

A/Gene: BCS4066

C/Superfamily: prenyl transferase A

Query Match 20.6%; Score 451.5; DB 2; Length 323;
Best Local Similarity 30.1%; Pred. No. 4.4e-24;
Matches 116; Conservative 51; Mismatches 132; Indels 87; Gaps 5;

QY 71 LVGPMSNLTNSIRSLSGHPSLDTVAKYVQSEGKHRLPLVLLMAQATEVAPKVGW 130

```
Db 9  LTAQDMAGVNAAILLEQLNSDVQLINQLGYIYVSGGKREIRPMIAVLARAV-----GY 61
QY 131 EKVEVPVNEGLAPPEVLNDKPNMNMRSGLTKDGEIEGQTSNIIASQRRLAIEITMI 190
Db 62  EGNAAHVTI-----AALIEFI 76
QY 191 HAASLLHDDVDIVASSTRNAPSNGNOAGNKMAIILAGDFLLGRASVALARLNPEVIELLA 250
Db 77 HTATLLHDDVVDESMDRGRKATANAAGNAASVLVGDFIYTRAFQMMTSLGSLKVLVMS 136
QY 251 TVIANLVEGEMQLKNTVDDAIEATATQETFDYYLQKTYLKTASLIASCRASALLGGAT 310
Db 137 EAVNVIAEGEVLQLMNVNDPDI-----TEENTYRVYISKYARLFEEAAQCSGILAGCT 189
QY 311 PEVADAAYAYGRNIGLAFQIVDDMLDYTVSATDLGKPGADLQGLATAPALFAWKHAE 370
Db 190 PEEKGLQDYGRYLGTAFLQIDLLDYNADGEQLGKNVGDDLNKGKPTPLLLHAMHH--- 246
QY 371 LGPMIKKFSDPGDVERARELVEKSDG-----LEKTRALAEYAAQKALD 414
Db 247 -----GTPEQAQMIRTAIEQGNRHLKPVLEAMNACGSLEWTRQRAEEADKAIA 297
QY 415 AIRTFPSPARKALEQLTDKVLTRSR 440
Db 298 ALQVLPDTPWREALIGLAHIAVQDR 323

RESULT 15
E85982
octaprenyl diphosphate synthase [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85982
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85982
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <STO>
A;Cross-references: CB:AE005174; NID:g12517802; PIDN:AGS8321.1; GSPDB:GN00145; UMGP:Z45
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ispB
C;Superfamily: prenyl transferase A

Query Match 20.6%; Score 451.5; DB 2; Length 323;
Best Local Similarity 30.1%; Pred. No. 4.4e-24;
Matches 116; Conservative 51; Mismatches 132; Indels 87; Gaps 5;

QY 71 LVGPMSNLTNSIRSLGSGHPSLDTVAKYVQSEGGKHIPELMVLMAQATEVAPKVQGW 130
Db 9  LTAQDMAGVNAAILLEQLNSDVQLINQLGYIYVSGGKREIRPMIAVLARAV-----GY 61
QY 131 EKVEVPVNEGLAPPEVLNDKPNMNMRSGLTKDGEIEGQTSNIIASQRRLAIEITMI 190
Db 62  EGNAAHVTI-----AALIEFI 76
QY 191 HAASLLHDDVDIVASSTRNAPSNGNOAGNKMAIILAGDFLLGRASVALARLNPEVIELLA 250
Db 77 HTATLLHDDVVDESMDRGRKATANAAGNAASVLVGDFIYTRAFQMMTSLGSLKVLVMS 136
QY 251 TVIANLVEGEMQLKNTVDDAIEATATQETFDYYLQKTYLKTASLIASCRASALLGGAT 310
Db 137 EAVNVIAEGEVLQLMNVNDPDI-----TEENTYRVYISKYARLFEEAAQCSGILAGCT 189
QY 311 PEVADAAYAYGRNIGLAFQIVDDMLDYTVSATDLGKPGADLQGLATAPALFAWKHAE 370
Db 190 PEEKGLQDYGRYLGTAFLQIDLLDYNADGEQLGKNVGDDLNKGKPTPLLLHAMHH--- 246
QY 371 LGPMIKKFSDPGDVERARELVEKSDG-----LEKTRALAEYAAQKALD 414
```

```
Db 247 -----GTPEQAQMIRTAIEQGNRHLKPVLEAMNACGSLEWTRQRAEEADKAIA 297
QY 415 AIRTFPSPARKALEQLTDKVLTRSR 440
Db 298 ALQVLPDTPWREALIGLAHIAVQDR 323

Search completed: January 16, 2003, 09:50:00
Job time : 39 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 16, 2003, 09:41:01 ; Search time 23 Seconds
(without alignments)
793.460 Million cell updates/sec

Title: US-09-830-111A-2
Perfect score: 2194
Sequence: 1 MASPALIRISRSIASLR.....ESPAKALEQLDKVLRSR 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023	46.6	473	1	COOL_YEAST
2	910.5	41.5	378	1	DPS_SCHPO
3	676.5	30.8	323	1	PREA_SYNY3
4	664.5	30.3	323	1	PREA_CVAPA
5	604.5	27.6	323	1	PREA_PORPU
6	571	26.0	323	1	PREA_CVACA
7	451.5	20.6	323	1	ISPB_ECOLI
8	445.5	20.3	329	1	ISPB_HAEIN
9	409	18.6	320	1	HEP2_BACST
10	406	18.5	348	1	HEP2_BACSU
11	365.5	16.7	325	1	IDSA_METHH
12	361.5	16.5	324	1	IDSA_METHM
13	309	14.1	359	1	GGPP_MYCTU
14	299	13.6	327	1	IDSA_METHA
15	292	13.3	332	1	GGPP_SULSO
16	284	12.9	371	1	GGPP_ARATH
17	274	12.5	366	1	GGPP_SINAL
18	265	12.1	369	1	GGPP_CAPAN
19	260	11.9	299	1	ISPA_ECOLI
20	260	11.9	330	1	GGPP_SULAC
21	256	11.7	357	1	GGPP_CATRO
22	247	11.3	297	1	ISPA_BACST
23	245	11.2	300	1	CRTE_CVAPA
24	210.5	9.6	272	1	ISPA_BACSU
25	210	9.6	291	1	ISPA_MICLU
26	206	9.4	307	1	CRTE_ERWHE
27	203.5	9.3	302	1	CRTE_PANAN
28	199.5	9.1	289	1	CRTE_RHOCA
29	198.5	9.0	288	1	CRTE_RHOSH
30	198.5	9.0	418	1	GGPP_GIBFU
31	191.5	8.7	295	1	ISPA_HAEIN
32	190	8.7	332	1	ISPA_BRAJA
33	184	8.4	332	1	ISPA_RHISN

34	178	8.1	428	1	GGPP_NEUCR	P24322 n geranylge
35	177	8.1	262	1	ISPA_AQUAE	O66952 aquifex ae
36	147	6.7	347	1	PPPS_NEUCR	Q92250 neurospora
37	143.5	6.5	300	1	GGPP_MOUSE	Q9wtm0 m geranylge
38	140.5	6.4	294	1	GGPP_BOVIN	P56966 b geranylge
39	139.5	6.4	300	1	GGPP_HUMAN	O95749 h geranylge
40	127.5	5.8	347	1	PPPS_GIBFU	Q92235 gibberella
41	121	5.5	739	1	OCT1_CHICK	P15143 gallus gall
42	119.5	5.4	347	1	PPPS_SCHPO	O14230 schizosacch
43	117	5.3	550	1	SYR_MYCTU	Q10609 mycobacteri
44	115.5	5.3	810	1	CLPC_BACSU	P37571 bacillus su
45	114.5	5.2	282	1	ISPA_BUCAI	P57537 buchnera ap

RESULT 1

ID	COOL_YEAST	STANDARD;	PRT;	473 AA.
AC	P18900;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hexaprenyl pyrophosphate synthetase, mitochondrial precursor			
DE	(EC 2.5.1.-) (HPS).			
GN	COQ1 OR YBR003W OR YBR0109.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=D273-10B/Al;			
RX	MEDLINE=90330660; PubMed=2198286;			
RA	Ashby M.N., Edwards P.A.			
RT	"Elucidation of the deficiency in two yeast coenzyme Q mutants.			
RT	Characterization of the structural gene encoding hexaprenyl			
RT	pyrophosphate synthetase."			
RL	J. Biol. Chem. 265:13157-13164(1990).			
RN	[2]			
RP	SEQUENCE OF i-285 FROM N.A.			
RC	STRAIN=S288c;			
RX	MEDLINE=94378721; PubMed=8091860;			
RA	Wolfe K.H., Lohan A.J.E.;			
RT	"Sequence around the centromere of Saccharomyces cerevisiae			
RT	chromosome II: similarity of CEN2 to CEN4."			
RL	Yeast 10:S41-S46(1994).			
RN	[3]			
RP	SEQUENCE OF 259-479 FROM N.A.			
RC	STRAIN=S288c;			
RA	Domdey H., Gassenhuber H., Obermaier B., Piravandi E.;			
RT	Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.			
CC	-I- FUNCTION: ASSEMBLY OF POLYISOPRENOID SIDE CHAINS. THE POLYPRENYL			
CC	ISOTENYL DIPHOSPHATE Q BIOSYNTHESIS CATALYZES THE FORMATION FROM			
CC	GENERALLY RANGING IN LENGTH OF BETWEEN 6 AND 10 ISOPRENE UNITS			
CC	DEPENDENT ON THE SPECIES.			
CC	-I- PATHWAY: Coenzyme Q biosynthesis; first step.			
CC	-I- SUBCELLULAR LOCATION: MAY BE PERIPHERALLY ASSOCIATED WITH THE			
CC	INNER MITOCHONDRIAL MEMBRANE.			
CC	-I- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J05547; AAA34686.1; -.			
DR	EMBL; Z26494; CAA81272.1; -.			

[illegible]

```

SP SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajendream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles T., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Moore P., O'Neill S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton T., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymptre B.,
RA Welljens I., Vansteelt E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzys K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Mambitt R., Punnelle B.,
RA Goffeau A., Cadieu E., Dreaon S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rocher M., Galliaridin C., Tallada V.A., Rayson A., Rhode G.,
RA Dugas R.R., Cuzado L., Jimenez J., Sanchez M., del Ray F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Snapakovski G.V., Useery D., Bartell B.G., Nurse P.;
RT "The genome sequence of Schistosoma haematophyllum pome."
RL Nature 415:871-880(2002).
CC -I- FUNCTION: SUPPLIES DECAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE
CC SIDE CHAIN OF THE ISOPRENOID UNIFORMS UBIQUINONE-10.
CC -I- SIMILARITY: BELONGS TO THE PPP/GGP SYNTHETASS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84311; BAA12314.1; -
DR EMBL; AL136498; CAB6154.1; -
DR InterPro; IPR000092; PolyPrenyl_syn.
DR Pfam; PF00348; PolyPrenyl_syn.1
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR Isoprene biosynthesis; ubiquinone biosynthesis; Transferase.
KW SEQUENCE 378 AA; 42046 MW; 84B3300F84EBB05E CRC64;
SQ
Query Match 41.5%; Score 910.5; DB 1; Length 378;
Best Local Similarity 51.1%; Pred. No.2.9e-55;
Matches 191; Conservative 60; Mismatches 96; Indels 27; Gaps 3;
QY 67 NPILNVGEEMSLTNRITSLGSGHPSLDYAKKYVVOSEGHIRPRLMTLMAQAEVAPK 126
DB 32 NASHIKKELRQISGIRIQMLNSNSEFLSECSKYTTIAQGQMRPSLVLLMSKATSLCHG 91
QY 127 VQGEKVEVFPNEGRLAPPEVLNDGNQPMNNMRSGPLTKDGBIEGQTSNIILASQRLAEI 186
DB 92 ID-----RSVVGKRTITDDDIRS-----FTQQLPSQRLAQI 125
QY 187 TEMIHASLLHDDVIDASETRRNAPSGNAFGKXMAIAGFLGRASVALARLNNPEVI 246
DB 126 TEMIHASLLHDDVIDAHANVRGSPSSNVAFGNRSSILAGNFILARASTAMARLNNPQT. 185
QY 247 ELLATVIANLVGEEMQAKNTVDDAIEKTAQETFDYLYLQKTYLKTASLIKSCASALL 306
DB 186 ELLATVIANLVGEEMQAKNTVDDAIEKTAQETFDYLYLQKTYLKTASLIKSCASATYL 244
QY 307 GGATEVVDAAVAVGRNIGLAFQIVDMLDVTYVASTDLGKAGADLGLGATAPALPMK 366
DB 245 GGCSFTVATTAAGEYRCIGTAFQMLDHDVLDYTSKDDTLGKAAGADLKGLATAPALPMK 304

```

QY 367 HHAELGPMIKRKESDGDVERARELVEKSDLEKTRALAEYQAOKALDAIRTFPSPARK 426
 Db 305 KYPELGAMIVNRPNHPSDIQARSLVECTDAIEQTITWAKKEYIKKAKDSLLCLPDSARK 364
 QY 427 ALQRLTKVLTRSK 440
 Db 365 ALPALADKVIITRKK 378

RESULT 3
 PREA SYN3
 ID PREA SYN3 STANDARD; PRT; 323 AA.
 AC P72580; P74758; O08042;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Prenyl transferase (EC 2.5.1.-).
 GN PREA OR SLR0611.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugiura M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 64% to 92% of the genome.";
 RL DNA Res. 2:153-166(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugiura M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- FUNCTION: POSSIBLE ROLE IN SYNTHESIS OF THE NONAPRENYL SIDE CHAIN
 CC OF PLASTOQUINONE OR IN SYNTHESIS OF OTHER PRENYL CHAINS SUCH AS
 CC UNDEKAPRENYL PYROPHOSPHATE.
 CC -!- SIMILARITY: BELONGS TO THE PFP/GGPP SYNTHETASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90899; BAA16579.2; --
 CC EMBL; D90917; BAA16579.2; JOINED.
 CC InterPro; IPR000092; Polyprenyl synt.
 CC Pfam; PF00348; polyprenyl synt; 1.
 CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
 CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; FALSE NEG.
 CC KW Photosynthesis; Isoprene biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE . 323 AA; 35725 MW; 42160CE6C165C685 CRC64;

Query Match 30.8%; Score 676.5; DB 1; Length 323;
 Best Local Similarity 40.6%; Pred. No. 2.4e-39;
 Matches 153; Conservative 63; Mismatches 102; Indels 59; Gaps 4;

QY 65 LINPLNLVGP---EMSNLTNSIRSLGSGHPSLDTVAKYVQSGKHIRPLMLVLLMAQAT 121
 Db 1 MISTTSFAPVDQDLRLTLNKLVLGARHPILGAAAEHLFEAGGKVRVPAIVLLVSRAT 60

QY 122 EVAPKVOGKEKVEVPVNEGLAPPEVLNDKNPDMNMRSPLTKDGEIGQTSNIIASOR 181
 Db 61 -----LLDQ-----ELTARHR 71
 QY 182 RLAEITEMIHAASLLHDDVIDASETRRNAPSGNQAEGNMAILAGDFLLGRASVALARLR 241
 Db 72 RLAEITEMIHTASLVHDDVDEADLRNVPVNSLFDNRVAVLAGDFLFAQSSWYLANLD 131
 QY 242 NPEVIELLATVIANLVEGEFMQLKNVTVDIAEATATQETFDYLYQTYLKTASLIAKSCR 301
 Db 132 NLEVVKLLSEVIRDFAEGBEILQSIINRFD-----TDTLETYLEKSYFKTASLIANSK 184
 QY 302 ASALLGGATPEVADAAAYAGRNLGLAFOIVDDMLDVTVSATDLGKPGADLQGLATAPA 361
 Db 185 AAGVLSADPRDVCDDHLYEYKGLGLAFOIVDDILDTSPTEVLGKPGAGDLSGNITAPA 244
 QY 362 LFAWKHHAELGPMIKRKESDGDVERARELVEKSDLEKTRALAEYQAOKALDAIRTFPE 421
 Db 245 LFAWKYPLGLKLIEREFAQAGDLEQALELVEGDDGIRRSRELAANQAQLARQHLVLEM 304
 QY 422 SPARKALEOLTDKVLTR 438
 Db 305 SAPRESLLELVYVLR 321

RESULT 4
 PREA CYAPA
 ID PREA CYAPA STANDARD; PRT; 323 AA.
 AC P31171;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Prenyl transferase (EC 2.5.1.-).
 GN PREA.
 OS Cyanophora paradoxa.
 OC Cyanophyta; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
 OX NCBI_TaxID=2762;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LB555 / Pringsheim;
 RX MEDLINE=91268060; PubMed=1711042;
 RA Michalowski C.B., Loeffelhardt W., Bohnert H.J.;
 RT "An ORF323 with homology to crtE, specifying prephytoene
 RT pyrophosphate dehydrogenase, is encoded by cyanelle DNA in the
 RT eukaryotic alga Cyanophora paradoxa.";
 RL J. Biol. Chem. 266:11866-11870(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LB555 / Pringsheim;
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
 RA Bryant D.A.;
 RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
 RL Plant Mol. Biol. Rep. 13:327-332(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LB555 / Pringsheim;
 RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
 RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
 RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
 RT the genetic complexity of a primitive plastid.";
 RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
 RL Schwemmler W. (eds.);
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
 RL (1997).
 RN [4]
 RP SEQUENCE OF 1-46 FROM N.A.
 RC STRAIN=LB555 / Pringsheim;
 RX MEDLINE=91117189; PubMed=2126059;
 RA Michalowski C.B., Pfanagl B., Loeffelhardt W., Bohnert H.J.;
 RT "The cyanelle S10 spc ribosomal protein gene operon from Cyanophora
 RT paradoxa.";

```

RL Mol. Gen. Genet. 224:222-231(1990).
CC - FUNCTION: POSSIBLE ROLE IN SYNTHESIS OF THE NONAPRENYL SIDE CHAIN
CC OF PLASTOQUINONE OR IN SYNTHESIS OF OTHER PRENYL CHAINS SUCH AS
CC UNDEKAPRENYL PYROPHOSPHATE.
CC - SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC - CAUTION: Was originally (Ref.1) called crte.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37111; AAA65472.1; -
CC EMBL; U30821; AAA81217.1; -
CC EMBL; M30487; AAA63631.1; -
CC PIR; A40433; A40433.
CC InterPro; IPR000092; Polyprenyl_synth.
CC Pfam; PF00348; polyprenyl_synth; 1.
CC PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
CC PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
CC Photosynthesis; Isoprene biosynthesis; Transferrase; Cyanelle.
SQ SEQUENCE 323 AA; 35919 MW; CB4802466342B09A CRC64;

Query Match 30.3%; Score 664.5; DB 1; Length 323;
Best Local Similarity 39.5%; Pred. No. 1.6e-38;
Matches 147; Conservative 69; Mismatches 97; Indels 59; Gaps 4;

QY 70 NLVDP---EKSNTLSNIRSLGSHPSLDTVAKYVQSEKRIPLMLVLMQAQTEVAPK 126
DB 6 NILAPVENELDLTKNKKLVGSGHPILLSASEHLFSAGGRPPATVLLISKA----- 59
QY 127 VQGEKVEVPEVNEGGLAPPEVLNDKNPDMNMRSGPLTKDGEIGQTSNIIASORLAEI 186
DB 60 -----TMEHEI-----ITSKRRLAEI 76
QY 187 TEMIHAASLLHDDVIDASETRRNAPSGNOAFGNKMAIILAGDFLLGRASVALARLNPEVI 246
DB 77 TEIHTASLVHDDILDESDDRGRIPVHSPGKIALIAGDFLPAQSSWYLANLESLEV 136
QY 247 ELATVIANLVEGEFMQKNTVDDAIEATATQETFDYLYOKTYIKTASLIASCRASALL 306
DB 137 KLISKIVITDPAEGEIR-----GLNQFVDTLLEBYLEKSPFYKTASLIASAASKAALL 189
QY 307 GGATPEVADAAYVAGNMLGLAFOIVDDMLDYVSATDLGKPPAGADLQGLATAPALPAPK 366
DB 190 SHVDLTIVANDLYNGRHLGLAFOIVDDILDTSTBELGKPSGSDLKGNLTAPVLPALF 249
QY 367 HHAELGPMIKRKFSPGDVERARELVEKSDGLEKTRALAEYAKALDAIRTPESPAPK 426
DB 250 QNSLILIGLQRFSEPPDFFETYLQIVETKAIKTRRLAMEHAQVVALQCEENLPSSSKE 309
QY 427 ALBQLTDKVLTR 438
DB 310 ALKLITKYLVER 321

```

```

RESULT 5
PREA_FORPU STANDARD; PRT; 323 AA.
ID PREA_FORPU
AC P51268;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE Prenyl transferase (EC 2.5.1.-).
GN Porphyrin purpurea.
OS Porphyrin purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyrin.
OX NCBI_TaxID=27877;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-Avonport;
RT Kelt M.E., Munholland U.;
RT "Complete nucleotide sequence of the porphyrin purpurea chloroplast
RT genome."
RT Plant Mol. Biol. Rep. 13:333-335(1995).
CC - FUNCTION: POSSIBLE ROLE IN SYNTHESIS OF THE NONAPRENYL SIDE CHAIN
CC OF PLASTOQUINONE OR IN SYNTHESIS OF OTHER PRENYL CHAINS SUCH AS
CC UNDEKAPRENYL PYROPHOSPHATE.
CC - SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U38804; AAC08154.1; -
CC InterPro; IPR000092; Polyprenyl_synth.
CC Pfam; PF00348; polyprenyl_synth; 1.
CC PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
CC PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
CC Photosynthesis; Isoprene biosynthesis; Transferrase; Chloroplast.
SQ SEQUENCE 323 AA; 35589 MW; 9FB02F4E2A2019B2 CRC64;

Query Match 27.6%; Score 604.5; DB 1; Length 323;
Best Local Similarity 36.1%; Pred. No. 2.1e-34;
Matches 135; Conservative 69; Mismatches 111; Indels 59; Gaps 3;

QY 65 LNPVLNVPENKSNITNIRSLGSHPSLDTVAKYVQSEKRIPLMLVLMQAQTEVA 124
DB 7 LPHPIE---KELYSVENHLKSVAGTRHPILYAAKHLEFAGCKRLRPAIVLLVAKST--- 60
QY 125 PVQGEKVEVPEVNEGGLAPPEVLNDKNPDMNMRSGPLTKDGEIGQTSNIIASORLAEI 184
DB 61 -----SQGEIKKQGRRLA 74
QY 185 EITEMIHAASLLHDDVIDASETRRNAPSGNOAFGNKMAIILAGDFLLGRASVALARLNPE 244
DB 75 ETEIHTASLVHDDILDESDDRGRIPVHSPGKIALIAGDFLPAQSSWYLANLESLEV 134
QY 245 VIELATVIANLVEGEFMQKNTVDDAIEATATQETFDYLYOKTYIKTASLIASCRASALL 304
DB 135 VVKVITKIVITDPAEGEIR-----GLNQFVDTLLEBYLEKSPFYKTASLIASAASKAALL 187
QY 305 LGGATPEVADAAYVAGNMLGLAFOIVDDMLDYVSATDLGKPPAGADLQGLATAPALPAPK 364
DB 188 MLNDLNSQWMDLVLYGKRMGLAFOIVDDILDTSTBELGKPSGSDLKGNLTAPVLPALF 247
QY 365 WGHHAELGPMIKRKFSPGDVERARELVEKSDGLEKTRALAEYAKALDAIRTPESPAPK 424
DB 248 LQGEKLDQGLQRFSEPPDFFETYLQIVETKAIKTRRLAMEHAQVVALQCEENLPSSSKE 307
QY 425 KRALBQLTDKVLTR 438
DB 308 VSLKELTHFTITR 321

```

```

RESULT 6
PREA_CYACA STANDARD; PRT; 323 AA.
ID PREA_CYACA
AC Q9TSL1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Prenyl transferase (EC 2.5.1.-).
GN Cyanidium caldarium.
OS Cyanidium caldarium.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.

```

```

OX NCBI_TaxID=2771;
RN SEQUENCE FROM N.A.
RC STRAIN=K-1;
RX MEDLINE=2049659; PubMed=11040290;
RA Gloeckner G., Rosenthal A., Valentin K.-U.;
RT "The structure and gene repertoire of an ancient red algal plastid genome.";
RL J. Mol. Evol. 51:382-390(2000).
CC -!- FUNCTION: POSSIBLE ROLE IN SYNTHESIS OF THE NONAPRENYL SIDE CHAIN OF PLASTOQUINONE OR IN SYNTHESIS OF OTHER PRENYL CHAINS SUCH AS UNDEKAPENYL PYROPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022186; AAF12896.1; -
DR InterPro; IPR000092; Polyprenyl synt.
DR Pfam; PF00348; polyprenyl synt.; 1
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
KW Photosynthesis; Isoprene biosynthesis; Transferase; Chloroplast.
SQ SEQUENCE 323 AA; 36052 MW; C9E7F23CBED3FC47 CRC64;

Query Match 26.0%; Score 571; DB 1; Length 323;
Best Local Similarity 35.8%; Pred. No. 4.1e-32;
Matches 133; Conservative 67; Mismatches 110; Indels 64; Gaps 4;

QY 69 LNLVGPMSNLTSNIRSLGSGHPSLDTVAKYVQSGKIRPLMVLMLMAQATEVAPKVQ 128
DB 8 LQSVKEDLLNIEQTLNKLKVNPNILSAAXHLLVSKKIRPAIVLLVAKI----- 60

QY 129 GWKEVVPVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNILASQRLAEITE 188
DB 61 -----DKN-----KKIKTSQORLAEBTE 78

QY 189 MIHAASLLHDDVIDASETRNAPSGNAGFKMMAILAGDFLGLPASVALARLNRPVIEL 248
DB 79 IHTATLLHDDVDSEIRRTGTSKWNTPGKNIAVFAFGDFLFAOSSWYLANINNVYKA 138

QY 249 LATVIANLVEGEFQMLKNTVDDAIEATATPTDY-----YLOKTYLKTASLIASKRASA 304
DB 139 ISKVITDLAEGELQ-----NLTPNTYYSIKVLEKSFNTASLIASCKSCC 187

QY 305 LGGATEVADAAYAGRNGLAFQIVDDMLDYTVSATDLGPKAGADLQGLATAPALFA 364
DB 188 LLSDFDQSLNSKFNFGKLNGLAFQIIDDITSSSTALGKMTTSLKGLNLTAPVLF 247

QY 365 WKHAEIOPMKRFPDGVVERRELVEKSDGLEKTRALAEVAQAALDAIRTFPSPA 424
DB 248 LTKNSKLFKIEREFCKSDISEAINIKETNAIEESFDLAEYHIEAINSIKDLPTSSE 307

QY 425 KALEQLTQDKVLR 438
DB 308 KDSLIEIAYDLLNR 321

RESULT 7
ISPB_ECOLI
ID ISPB_ECOLI STANDARD; PRT; 323 AA.
AC F19641;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Octaprenyl-diphosphate synthase (EC 2.5.1.-) (Octaprenyl pyrophosphate synthetase) (OPP synthetase).
GN ISPB OR CEL OR B3187.

```

```

OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94146411; PubMed=8312607;
RA Jeong J.H., Kitakawa M.S., Isono S., Isono K.;
RT "Cloning and nucleotide sequencing of the genes, rpiU and rpmA, for RT ribosomal proteins L21 and L27 of Escherichia coli.";
RL DNA Seq. 4:59-67(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 196-323 FROM N.A.
RX MEDLINE=89359178; PubMed=2670911;
RA Choi Y.-L., Nishida T., Kawamukai M., Utsumi R., Sakai H., Komano T.;
RT "Cloning and sequencing of an Escherichia coli gene, nlp, highly RT homologous to the ner genes of bacteriophages Mu and D108.";
RL J. Bacteriol. 171:5222-5225(1989).
RN [4]
RP FUNCTION.
RC STRAIN=K12 / JM109;
RX MEDLINE=94311902; PubMed=8037730;
RA Asai K.-I., Fujisaki S., Nishimura Y., Nishino T., Okada K., RA Nakagawa T., Kawamukai M., Matsuda H.;
RT "The identification of Escherichia coli ispB (cel) gene encoding the RT octaprenyl diphosphate synthase.";
RL Biochem. Biophys. Res. Commun. 202:340-345(1994).
CC -!- FUNCTION: SUPPLIES OCTAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE SIDE CHAIN OF THE ISOPRENOID QUINONES UBIQUINONE AND MENAQUINONE.
CC -!- SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13267; -; NOT ANNOTATED_CDS.
DR EMBL; U18997; AAA57988.1; -
DR EMBL; A5000399; AAC76219.1; -
DR EMBL; X68873; CAA48735.1; ALT_SEQ.
DR PIR; PV0010; PV0010.
DR EcoGene; EG10017; ispB.
DR InterPro; IPR000092; Polyprenyl synt.
DR Pfam; PF00348; polyprenyl synt.; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
KW Isoprene biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 323 AA; 35217 MW; 08AD7AE7AC230EF8 CRC64;

Query Match 20.6%; Score 451.5; DB 1; Length 323;
Best Local Similarity 31.0%; Pred. No. 6.3e-24;
Matches 116; Conservative 52; Mismatches 143; Indels 63; Gaps 4;

QY 71 LVGPMSNLTSNIRSLGSGHPSLDTVAKYVQSGKIRPLMVLMLMAQATEVAPKVQGW 130
DB 9 LTAQDMAGVNAALIEQLNSDVQLINQLNGYIVSGGKIRIRPMIAVLAARAV-----GY 61

QY 131 EKVEVPVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNILASORRLAEITE 190
DB 131 EKVEVPVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNILASORRLAEITE 190

```

```

Db      62 EGNAAHTI-----ALIEFI 76
QY      191 HAASLLHDDVIDASETRRNAPSGNQAQFNKAILAGDFLLGRASVALARLNPEVIELLA 250
      77 HTATLLHDDVDVDESDMRGRATANAAGNAAASYLVGFITRAFOVMTSISGSIKVLEWVS 136
QY      251 TVIANLVGEFMOQKNVDAIEATATQEFEDYLOKTYIKTSLAKSCRASALLGGAT 310
      137 EAVNVIAEGEVLOLMANNDPDI-----TEENVRYIYKTAFLFEAAQCGGILAGCT 189
QY      311 PEVADAAYAGRNGLAFOIYVDMLDYTVASATDLGKPGADLQGLATAPALPAMKHAE 370
      190 PEEKGLQDGRYIGTAFQILDLDLDYNADGEOGKNVGDIDNKGKPTLPILHAMHGT 249
QY      371 LGMPIKRSFDPGD---VERARELVKSDGLEKTRALAEVYOKALDAIRTPESPARK 426
      250 EOKMIRTAIEQGNRHLLEPVLNANACSLLEWTRORAEEDAKAIALQVLPDTPWRE 309
QY      427 ALQOLDKVLTISR 440
Db      310 ALIGLAHIAVORDR 323

RESULT 8
ISPB_HAEIN
ID ISPB_HAEIN STANDARD; PRT; 329 AA.
AC P44916;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Octaprenyl-diphosphate synthase (EC 2.5.1.-) (Octaprenyl pyrophosphate
synthetase) (OPP synthetase).
GN ISPB OR H10881.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utecherback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512 (1995).
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429 (2000).
CC -1- FUNCTION: SUPPLIES OCTAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE
CC SIDE CHAIN OF THE ISOPRENOID QUINONES UBIQUINONE AND MENAQUINONE
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR      EMBL: U32770; AAC22540.1; -.
DR      TIGR: H10881; -.
DR      InterPro: IPR000092; Polyprenyl_synth.
DR      Pfam: PF00348; Polyprenyl_synth; 1.
DR      PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
DR      PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
DR      K01996; biosynthesis; Transferrase; Complete proteome.
SQ      SEQUENCE 329 AA; 35911 MW; 65DBCA4CCDAD072BE4 CRC64;

Query Match      20.3%; Score 445.5; DB 1; Length 329;
Best local similarity 32.5%; Pred. No. 1,7e-23;
Matches 119; Conservative 49; Mismatches 133; Indels 65; Gaps 6;

QY      71 LVGPENSLTSTIRSLSCGHPSLDTVAKYVYVSGEKIRPLMVLMAQAEVAPKQGW 130
      15 LADPMQKVNQNIILQINSVPLIGQLGFYIVQGGKRIPLVIAVARSU----- 65
Db      131 EKVEVPVNEGLAPPEVLNDKNPDMNMNRSGLPTKDGIEQTSNIIASQRLAITEMI 190
      66 -----GFESNSIITCAT-----PVEFI 82
QY      191 HAASLLHDDVIDASETRRNAPSGNQAQFNKAILAGDFLLGRASVALARLNPEVIELLA 250
      83 HTASLLHDDVDVDESDMRGRATANAAGNAAASYLVGFITRAFOVLAQLESILKILSI 142
QY      251 TVIANLVGEFMOQKNVDAIEATATQEFEDYLOKTYIKTSLAKSCRASALLGGA 309
      143 DATNVLAEGEVQOLMANVNDPETERSEAN-----YMRVYKTAFLFEVAAQAAIYVAGG 194
QY      310 PEVADAAYAGRNGLAFOIYVDMLDYTVASATDLGKPGADLQGLATAPALPAMKH-H 368
      195 TTAQEKALQDGRYIGTAFQILDVDLDYSANTQALGKNVGDIDNKGKPTLPILHAMHGN 254
QY      369 AELGPIKRSFDPGDVERARE---LVKESDGLEKTRALAEVYOKALDAIRTPESPARK 425
      255 AQQAALIREALIEQGGKRAIDEVALIMTEHKSIDYANMRKAKEAQAQVDAIEILPESBYK 314
Db      426 KALBOL 431
      315 QALISL 320

RESULT 9
HEBP2_BACST
ID HEBP2_BACST STANDARD; PRT; 320 AA.
AC P55785;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hepcaprenyl diphosphate synthase component II (EC 2.5.1.30) (HEPPP
DE synthase subunit 2).
GN HEPY OR HEP5-2.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OC NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 10149;
RX MEDLINE=95355463; PubMed=7629164;
RA Kojima-Takeshita A., Koyama T., Obara S., Ogura K.;
RA "Molecular cloning and nucleotide sequences of the genes for two
RT essential proteins constituting a novel enzyme system for heptaprenyl
RT diphosphate synthesis.";
RL J. Biol. Chem. 270:18396-18400 (1995).
CC -1- FUNCTION: SUPPLIES HEPYAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE
CC SIDE CHAIN OF THE ISOPRENOID QUINONE MENAQUINONE-7 (MQ-7).
CC -1- CATALYTIC ACTIVITY: All-trans-hexaprenyl diphosphate + isopentenyl
CC diphosphate = diphosphate + all-trans-heptaprenyl diphosphate.
CC -1- SUBUNIT: HETERODIMER OF COMPONENT I AND II.
CC -1- SIMILARITY: BELONGS TO THE PPP/GGPP SYNTHETASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D49976; BAA08725.1; --
CC InterPro; IPR000092; Polyrenyl synt.
CC Pfam; PF00348; polyrenyl synt; 1.
CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
CC Isoprene biosynthesis; Transferase.
CC SEQUENCE 320 AA; 35808 MW; A5186630E6B88F46 CRC64;

Query Match 18.6%; Score 409; DB 1; Length 320;
Best Local Similarity 28.5%; Pred. No. 5e-21;
Matches 106; Conservative 61; Mismatches 127; Indels 78; Gaps 6;

QY 70 NLVGPENSLTNSIRLLSGHPSLDTVAKYVQSEKGRHPLMLVILMAQATEVAPKVQ 129

Db 8 SFUSDDLAEEELERAVQSEYGLGALHLLQAGKIRPVEVLLAARF----- 58

QY 130 WEKVVEVPVNEGLAPPEVLNDKNDPMNMRSGPLTKDGEIGQTSNILASORRLAEITEM 189

Db 59 -----GQYD--LERMKHVAVALEL 75

QY 190 IHAASLLHDDVIDASSTRNAPSGNQAFGNKMAILAGDFLLGRASVALARLNPEVIELL 249

Db 76 IHMASLVHDDVIDDADLRGRPTIKAKWSNRFAMVTGDFLFARSLERMAELGNPRAHQVL 135

QY 250 ATVIANLVEGEFMQLKNTVDDAIEATATQETFDY----YLOKTYLKTASLIASCRASAL 305

Db 136 AKTIVEVCRGEIEQIKD-----KYRFDQPLRYLRIRRKRTALLIAASCQLGAL 184

QY 306 LGGATPEVADAAYAGRNGLGAFQIVDDMLDYTSATDLGKPGADLQGLATAPALFAW 365

Db 185 AAGAPEIVKRLVWFGVHGVMSFQITDILFTGTEQLGKPGADLQGNVTLPVLVYAL 244

QY 366 KHH-----AELGPMIKRKFSDPDGVARARELVKSGLEKTRALABEYAKALDAIRTF 419

Db 245 SDERVKAIAAVGPE-----TDVAEMAIVSAIKRTDAIERSYALSRYLDKALHLLDGL 299

QY 420 PESPAKALEQL 431

Db 300 PMNEARGLLRDL 311

RESULT 10
HEP2_BACSU STANDARD; PRT; 348 AA.
AC P31114;
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heptaprenyl diphosphate synthase component II (EC 2.5.1.30) (HEPPP
DE synthase subunit 2) (Spore germination protein C3).
GN HEP2 OR HEPB OR GERCC OR GERCC3.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=168;
RA Henner D.J.;
RT "Sequence of Bacillus subtilis dbpA, mtr(A,B), gerC(1-3), ndk, cheR,
RT aro(B,E,F,H), trp(A-F), hisH, and tyrA genes."
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]

CHARACTERIZATION OF GERCC LOCUS.
RX MEDLINE=91037938; PubMed=2121900;
RA Yazdi M.A.; Moir A.;
RT "Characterization and cloning of the gerC locus of Bacillus subtilis
RT 168."

RL J. Gen. Microbiol. 136:1335-1342(1990).
RN [3]
RP FUNCTION.

RX MEDLINE=98386502; PubMed=9720033;
RA Leatherbarrow A.J.H., Yazdi M.A., Curson J.P., Moir A.;
RT "The gerC locus of Bacillus subtilis, required for menaquinone
RT biosynthesis, is concerned only indirectly with spore germination.";

RL Microbiology 144:2125-2130(1998).

CC -!- FUNCTION: SUPPLIES HEPTAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE
CC SIDE CHAIN OF THE ISOPRENOID QUINONE MENAQUINONE-7 (MQ-7).

CC -!- CATALYTIC ACTIVITY: All-trans-hexaprenyl diphosphate + isopentenyl
CC diphosphate = diphosphate + all-trans-heptaprenyl diphosphate.

CC -!- SUBUNIT: HETERODIMER OF COMPONENT I AND II.

CC -!- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M80245; AAA20856.1; --
DR EMBL; Z99115; CABI4190.1; --
DR Subtilist; BG10281; hept.

DR InterPro; IPR000092; Polyrenyl synt.
DR Pfam; PF00348; polyrenyl synt; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR Sporulation; Isoprene biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 348 AA; 39515 MW; OFF9C9199F04BE CRC64;

Query Match 18.5%; Score 406; DB 1; Length 348;
Best Local Similarity 27.7%; Pred. No. 9.1e-21;
Matches 101; Conservative 69; Mismatches 125; Indels 70; Gaps 6;

QY 70 NLVGPENSLTNSIRLLSGHPSLDTVAKYVQSEKGRHPLMLVILMAQATEVAPKVQ 129

Db 36 SFLNDDIDVIERELEQTVRSYDPLLEAGLHLLQAGKIRPVEVLL----- 82

QY 130 WEKVVEVPVNEGLAPPEVLNDKNDPMNMRSGPLTKDGEIGQTSNILASORRLAEITEM 189

Db 83 -----SGMF--GDYD-----INIKYVAVTLEM 103

QY 190 IHAASLLHDDVIDASSTRNAPSGNQAFGNKMAILAGDFLLGRASVALARLNPEVIELL 249

Db 104 IHMASLVHDDVIDDADLRGRPTIKAKWSNRFAMVTGDFLFARSLERMAELGNPRAHQVL 163

QY 250 ATVIANLVEGEFMQLKNTVDDAIEATATQETFDYLYLOKTYLKTASLIASCRASALGGA 309

Db 164 SQTIVEVCLGEIEQIKDKYN-----MEQNLRTRYLRIRRKRTALLIAVSCQLGAISGA 216

QY 310 TPEVADAAYAGRNGLGAFQIVDDMLDYTSATDLGKPGADLQGLATAPALFAWKHHA 369

Db 217 DEKHKALYVFGYVGVMSYQIIDIILFTSTEBELGKPGVGGDLQGNVTLPVLVYALKN-- 274

QY 370 ELGPMIKRKF-----SDPDGVARARELVKSGLEKTRALABEYAKALDAIRTFPESP 423

Db 275 ---PALKNQKLINSETTQQLPEIEIKKTDIAEASMAVSEMYLOKAFQKLTLPGR 331

QY 424 ARKAL 428

Db 332 ARSSL 336

RESULT 11
IDSA_METH STANDARD; PRT; 325 AA.
ID IDSA_METH
AC O26156;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Bifunctional short chain isoprenyl diphosphate synthase [includes:
DE Farnesyl pyrophosphate synthetase (EC 2.5.1.1) (FPP synthetase)]
DE (Dimethylallyltransferase); Geranyltransferase (EC 2.5.1.10)].
GN IDSA OR MTH50.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
NC NCBI_TaxID=187420;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delongchery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaeh R., Blakey D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pochler B., Qiu D.,
RA Spadefora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-T., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranyl diphosphate.
CC -1- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans-farnesyl diphosphate.
CC -1- COFACTOR: REQUIRES MAGNESIUM.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000797; AAB84557.1; -.
DR InterPro; IPR000092; Polyprenyl synt.
DR Pfam; PF00348; Polyprenyl synt. 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW Lipid synthesis; Isoprene biosynthesis; Transferase; Magnesium;
KW Multifunctional enzyme; Complete proteome.
SQ SEQUENCE 325 AA; 35485 MW; BCF3A285A207916E CRC64;

Query Match 16.7%; Score 365.5; DB 1; Length 325;
Best Local Similarity 29.3%; Pred. No. 4,9e-18;
Matches 109; Conservative 61; Mismatches 133; Indels 69; Gaps 9;

OY 75 EMSLTSNLSLGSQHP-SLDTVAKYVQSEGHIRPLMLWMAQTEVAAPKVGWEKV 133
DB 13 EMD--ERIESISDITPERLIRASEHLITAGGKIRPSLALLSSEAVGDP----- 62
OY 134 VEVVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNIIASQRRLAETEMTHAA 193
DB 63 -----GDAAG-----VAAAEILHTF 78
OY 194 SLHDDVIDASETRRNAPSGNQAFNGKMAILAGDFLLGRASVALARLRNEVLI-ELLATV 252
DB 79 SLHDDIMDDDEIRRGEPAAHVLMGEPMAILAGDVLFPSKAFVAVIRNGDEMKELAAV 138
OY 253 IANLV---EGEFMOLKNTVDADATATQETPRVYLQKTYLKTASLAKSCRAISALLGCA 309
DB 139 VDSGVKICEGQALDM--GFEERLDVTEE-----YMEMIYKTRALLAAATKAGALMG 191
OY 310 TPEVADAAYVYGRNLGLAFOIVDDMLDYVSATDLGRPAGADQLGLATAPALFAWKHNA 369
DB 192 SPOEIALBEDYGCIGLAFQIHDDYLDVDSDESLGKPVSDIAEGKMTLMVVKALERAS 251
OY 370 ELGPMIKRKSDFGD---VERARELVKSDGLEKTRALAEYVOKALDAIRTPESPARK 426

```

```

DB 252 EKDRERLSILSGGEKLVAAEAIIEFYRGATEYAHAVLDHVMARERLEESDARE 311
OY 427 ALEQLDKVLT 438
DB 312 ALAMTADPFLER 323

RESULT 12
IDSA_METTMM STANDARD; PRT; 324 AA.
AC 053479;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bifunctional short chain isoprenyl diphosphate synthase [includes:
DE Farnesyl pyrophosphate synthetase (EC 2.5.1.1) (FPP synthetase)
DE (Dimethylallyltransferase); Geranyltransferase (EC 2.5.1.10)].
GN IDSA.
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
OC Archaea; Euryarchaeota; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
NC NCBI_TaxID=79929;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95070145; PubMed=7979381;
RA Chen A., Poulter C.D.;
RT "Isolation and characterization of idsa: the gene for the short chain
RT isoprenyl diphosphate synthase from Methanobacterium
RT thermoautotrophicum.";
RL Arch. Biochem. Biophys. 314:399-404(1994).
CC -1- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranyl diphosphate.
CC -1- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans-farnesyl diphosphate.
CC -1- COFACTOR: REQUIRES MAGNESIUM.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S75695; AAB32421.1; -.
DR InterPro; IPR000092; Polyprenyl synt.
DR Pfam; PF00348; Polyprenyl synt. 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW Lipid synthesis; Isoprene biosynthesis; Transferase; Magnesium;
KW Multifunctional enzyme.
FT INTRMET 0
SQ SEQUENCE 324 AA; 35505 MW; 88558A8ADDF5F66C1 CRC64;

Query Match 16.5%; Score 361.5; DB 1; Length 324;
Best Local Similarity 31.4%; Pred. No. 9,2e-18;
Matches 107; Conservative 58; Mismatches 134; Indels 42; Gaps 8;

OY 121 TEVAPKVGWEKVEVPVNEGLAPPEVLNDKNPDMNMRSGPLTKDGEIEGQTSNIIAS- 179
DB 1 TEVLILRKYSVADKRI-----MECISDITPPTLLKASHLITAGGKIRPSLALLSC 54
OY 180 -----QRLAETEMTHAAASLHDDVDADASETRRNAPSGNQAFNGKMAILAGDFLL 230
DB 55 EAVGNPEDAAGVAAAIELIHTFSLIHDDIMDDEMRGEPSVAVINGEPMAILAGGVLF 114
OY 231 GRASVALARLRNPE-VIELATVIANLV---EGEFMOLKNTVDADATATQETPRVYLQ 286
DB 115 SKAFPAVAVIRNGDSERVKALAVVDSGVKICEGQALDM--GFEERLDVTE-----DEYME 167

```


QY 287 KTYLKTASLAKSCRSALGGATPEVADAAAYAGRNGLAFQIVDDMLDYTSATDLGK 346
 DB 168 MIYKTAALIAAATKAAGMGASEREVEALDYGKFIAGLAFQIHDDYLDVWSDEESLGK 227
 QY 347 PAGADLQGLATAPALFAWKHAE-----LGPWKRRKFPDGVVERARELVEKSDG 397
 DB 228 PVGSDIAEGKWTLMVVKALAESEDERLISILGS-----GDEGSVAEAEIFERYGA 281
 QY 398 LEKTRALAEYAQKALDAIRTFPSPARKALEQUTDKVLTR 438
 DB 282 TOVAHEVALDVVRMAKERLEILEDSDARDALMRADFVLER 322

RESULT 13

GGPP_MYCTU STANDARD; PRT; 359 AA.
 ID GGPP_MYCTU
 AC Q50737;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable geranylgeranyl pyrophosphate synthetase (GGPP synthetase)
 DE (GGPPAS) (Geranylgeranyl diphosphate synthase) [Includes:
 DE Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase
 DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
 GN RV3398C OR MT3506 OR MTC78.30.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98235987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -I- FUNCTION: CATALYZES THE TRANS-ADDITION OF THE THREE MOLECULES OF
 CC IPP ONTO DMAPP TO FORM GERANYLGERANYL PYROPHOSPHATE WHICH IS A
 CC PRECURSOR OF THE ETHER-LINKED LIPIDS.
 CC -I- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranyl diphosphate.
 CC -I- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
 CC = diphosphate + trans-trans-farnesyl diphosphate.
 CC -I- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranylgeranyl diphosphate.
 CC -I- PATHWAY: BIOSYNTHESIS OF MEMBRANE ETHER-LINKED LIPIDS.
 CC -I- SIMILARITY: BELONGS TO THE FPP/GGPP SYNTHETASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; Z717165; BAB01025.1; -
 DR EMBL; AE007156; AAK47843.1; -
 DR TIGR; MT3506; -
 DR TubercuList; RV3398C; -
 DR InterPro; IPR000092; Polyprenyl_synt.
 DR Pfam; PF00348; polyprenyl_synt; 1.
 DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
 DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
 KW Hypothetical protein; Lipid synthesis; Isoprene biosynthesis;
 KW Transferase; Complete proteome.
 SQ SEQUENCE 359 AA; 38852 MW; F59F229B19432E6E CRC64;

Query Match 14.1%; Score 309; DB 1; Length 359;
 Best Local Similarity 25.2%; Pred. No. 4.2e-14;
 Matches 102; Conservative 57; Mismatches 162; Indels 84; Gaps 9;
 QY 58 EDPNPNPLNPLNLVGPENSLTNSIRSLGSGHPSLDTVAKYVV-----QSEG 106
 DB 13 QPDSRMTTRTLPLVGLAHELITFTLRQADRLDPHMRPVVSYHLGWSDERGRPVNNCG 72
 QY 107 KHIRPLMWLMAOATEVAPKQGWKVEVPVNEGLAPPEVLNDKPNMNMRSGLPTKD 166
 DB 73 KAIRPALVFVAEAGADP-----H 92
 QY 167 GETEGQTSNIIASQRRLAEITEMHAASLIHDDVIDASETRRNAPSQNAQFNKMAILAG 226
 DB 93 SAIPGAVS-----VELVHNFSLVHDDLMDRDEHRRHPTVWALWGDMALLAG 140
 QY 227 DFLGRASVALARLNPEV---IELLATVIANLVEGEFMQLKNTVDDAIEATATQETFDY 283
 DB 141 DAMLSLAHEVLDDCDSPHVGAALRAISEATRELIRGQ-----AADTAFE-SRTDVALDE 193
 QY 284 YLQKTYLKTASLAKSCRSALGGATPEVADAAAYAGRNGLAFQIVDDMLDYTSATD 343
 DB 194 CLKWAEKTAALMAASAEVGAALLAGAPRSREALVAYGRHIGLAFQIVDDLLGIGWGRPEI 253
 QY 344 LGKPAGADLQGLATAPALFAWKHAEGLPMIKRKFSD-----PGDVERARELVEKS 395
 DB 254 TGFVPVSDLSRKKTLPTVWTVAGSGSAGRLLAANLVDETGSGTASDDELAALAEIECG 313
 QY 396 DGLEKTRALAEYAQKALDAIRT--FPSPARKALEQUTDKVLTR 438
 DB 314 GGRWASAEARRHVTQIGIDMWARIGIPDRPAE-LQDLAHYIVDR 357

RESULT 14

IDSA_METJA
 ID IDSA_METJA STANDARD; PRT; 327 AA.
 AC Q58270;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bifunctional short chain isoprenyl diphosphate synthase [Includes:
 DE Farnesyl pyrophosphate synthetase (EC 2.5.1.1) (FPP synthetase)
 DE (Dimethylallyltransferase); Geranyltransferase (EC 2.5.1.10)].
 GN IDSA OR MJ0860.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,


```

Db 138 AFSIFITSIIEISGQMDMSFENRLD-----VTEEE---YIQMIKGTAMLFSCSAAL 188
Qy 303 SALLGGATPEVADAAYAGRNLGLAFOIVDDMLDYTVSATDLGKPGAGADLQLGLATAPAL 362
Db 189 GGIINKANDDVKKLTGYGLNLGISFQIVDDILGIGDEKELGKPIYSDIREGKKTILVI 248
Qy 363 FAWKHAE-----LGPWIKRKFSDPGDERARELVEKSDGLEKTRALAEYAAOKALDAIR 417
Db 249 KTLSEATEDEKKILVSTLGNKEAKKEDLERASEIIRK-HSLQYAYDLAKKYSDLAIEMLR 307
Qy 418 TFPES--PARKALEQLTDKVLTRSR 440
Db 308 EIPVSNKTAEKALKYLAQFTIQRK 332

```

Search completed: January 16, 2003, 09:49:01
 Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 16, 2003, 08:53:11 ; Search time 67 Seconds
(without alignments)
1353.145 Million cell updates/sec

Title: US-09-830-111A-2
Perfect score: 2194
Sequence: 1 MSPALRISRSRSIASLR.....ESPAKALQLDKVLTRSR 440

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	823.5	37.5	356	3 P78914	P78914 schizosacch
2	742	33.8	430	5 Q9V923	Q9V923 drosophila
3	734	33.5	416	10 Q9FET8	Q9FET8 quercus rob
4	710.5	32.4	422	10 Q9F789	Q9F789 arabidopsis
5	706	32.2	423	10 Q8RWM1	Q8RWM1 arabidopsis
6	690	31.4	336	11 Q9WU69	Q9WU69 mus musculus
7	683	31.1	376	4 Q9Y2W5	Q9Y2W5 homo sapien
8	678	30.9	321	10 Q9FSW8	Q9FSW8 citrus sine
9	661.5	30.2	323	16 Q8Z0J7	Q8Z0J7 anabaena sp
10	641.5	29.2	393	5 P91093	P91093 caenorhabdi
11	625.5	28.5	245	5 Q8SY08	Q8SY08 drosophila
12	585	26.7	406	10 Q8S948	Q8S948 arabidopsis
13	583.5	26.6	390	10 Q8SYN0	Q8SYN0 arabidopsis
14	574.5	26.2	379	10 Q9SHG4	Q9SHG4 arabidopsis
15	561	25.6	297	10 Q64684	Q64684 arabidopsis
16	550	25.1	363	5 Q964Q8	Q964Q8 trypanosoma

17	455.5	20.8	322	16 Q8XVK7	Q8XVK7 ralistonia s
18	451.5	20.6	323	16 Q8XFR7	Q8XFR7 salmonella
19	451.5	20.6	323	16 Q8X9K0	Q8X9K0 escherichia
20	435	19.8	338	16 Q92RMO	Q92RMO rhizobium m
21	434	19.8	322	2 Q9RBZ8	Q9RBZ8 pseudomonas
22	424.5	19.3	348	16 Q9KUT1	Q9KUT1 vibrio chol
23	420.5	19.2	362	16 Q8YFI8	Q8YFI8 brucella me
24	418.5	19.1	323	16 Q8ZBAS	Q8ZBAS yersinia pe
25	418.5	19.1	358	16 Q8UHP6	Q8UHP6 agrobacteri
26	416	19.0	336	16 Q9F2X8	Q9F2X8 streptomyce
27	415	18.9	338	16 Q9AG11	Q9AG11 caulobacter
28	411.5	18.8	320	16 Q9RVU0	Q9RVU0 deinococcus
29	407.5	18.6	448	5 Q9VP87	Q9VP87 drosophila
30	405	18.5	322	16 Q9HVL5	Q9HVL5 pseudomonas
31	404.5	18.4	325	2 Q24743	Q24743 rhodobacter
32	402.5	18.3	330	16 Q9CNS9	Q9CNS9 pasteurella
33	402.5	18.3	335	16 Q66428	Q66428 mycobacteri
34	395	18.0	315	2 Q82832	Q82832 gluconobact
35	395	18.0	332	16 Q9PDI8	Q9PDI8 xylella fas
36	381.5	17.4	399	4 Q9NR58	Q9NR58 homo sapien
37	378	17.2	324	16 Q9UQV2	Q9UQV2 neisseria m
38	377.5	17.2	338	16 Q986C2	Q986C2 rhizobium 1
39	377	17.2	321	16 Q92A78	Q92A78 listeria in
40	368	16.8	319	16 Q9CFX2	Q9CFX2 lactococcus
41	368	16.8	330	16 Q9CBA5	Q9CBA5 mycobacteri
42	368	16.8	538	5 Q96130	Q96130 plasmodium
43	365.5	16.7	340	17 Q8U1V3	Q8U1V3 pyrococcus
44	363.5	16.6	322	16 Q8R7D2	Q8R7D2 thermoplasma
45	362.5	16.5	319	16 Q99U20	Q99U20 staphylococ

ALIGNMENTS

RESULT 1

P78914	PRELIMINARY;	PRT;	356 AA.
ID P78914			
AC P78914			
DT 01-MAY-1997	(Tremblrel. 03, Created)		
DT 01-JAN-1998	(Tremblrel. 05, Last sequence update)		
DT 01-DEC-2001	(Tremblrel. 19, Last annotation update)		
DE mRNA	, partial cds, clone: SV 1754.		
OS Schizosaccharomyces pombe (Fission yeast).			
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC Schizosaccharomycetes.			
OX NCBI_TaxID=4896;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=PR745;			
RX MEDLINE=98162722; PubMed=9501991;			
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;			
RT "Identification of open reading frames in Schizosaccharomyces pombe			
RL DNA Res. 4:363-369(1997).			
DR EMBL; D89265; BAA13926.1; .			
DR InterPro; IPR000092; Polypreryl synt.			
DR Pfam; PF00348; polypreryl synt. 1.			
DR PROSITE; PS00723; POLYPRERYL_SYNTHET_1; 1.			
DR PROSITE; PS00444; POLYPRERYL_SYNTHET_2; 1.			
SQ SEQUENCE 356 AA; 39692 MW; 833865A1780715AE CRC64;			

Query Match 37.5%; Score 823.5; DB 3; Length 356;

Best Local Similarity 50.0%; Pred. No. 7.6e-49;

Matches 175; Conservative 56; Mismatches 92; Indels 27; Gaps 4;

QY 67 NPLNLVGPMSNLTSTNIRSLGSHPSLDTVAKYVQSEGHKIRPLMLVLAQATEVAPK 126

Db 32 NASHLIKNELEQISPGIROMLNSNPELECSKYTTIAQCKMRPSLVLLMSKATSL--- 88

QY 127 VQWEKVVVEVFNVEGLAPPEVLNDKNDPMNMRSGPLTKDGEIEGQTSNILLASORLAEI 186

Db 89 FHGIDR-----SVVGDKYTDGDLRS-----FSTGQILPSQLRLAQI 125

Oy	187	TEMIHASILHDDVDVIDASETRRNAPSQNGFGNMALADPFLIGASVALRLNREVI	246
Dd	126	TEMIHASILHDDVDVIDDHANVRGSSPSNVAFGNRRSILAGNPFLIARASTMARLRNPQT	185
Oy	247	ELLATVIANLVEGEFMOLKNTVDVAIEATRTAQTFEDYYLOKTYLKTRSLIASCRSALL	306
Dd	186	ELLATVIAIDLVRGEFLOTUKNTMDES-SLEIKQSINFYDIKEKSPFKTASLISSCKRSTLL	244
Oy	307	GGAPFEVDAAVAAVGRNLGLAFOIVDDMLPYWVSATDGLGPKAGADLOLGATAPALFAWK	366
Dd	245	GQCSTVTATAGERCGICGTAFOLMDVDLDITTSKDPTLGAAGAQLDKLGATIPVLFAWK	304
Oy	367	HHAELGPMTIKRKFSDPGDVERBAELVEKSQGLEKTRALAEYAOKALDI	416
Dd	305	KYPELGAMTVNRFNHPSPIDQARSRLVECSDAIBEQPIWMAIEYIKKPPSV	354
<hr/>			
RESULT 2			
Ov9v9z3		PRELIMINARY;	PRT; 430 AA.
AC	09v9z3:		
Dt	01-MAY-2000 (TREMBLrel_13, Created)		
Dt	01-MAY-2000 (TREMBLrel_13, Last sequence update)		
Dt	01-JUN-2001 (TREMBLrel_17, Last annotation update)		
De	CG3684 protein.		
GN	CG3684.		
OS	Drosophila melanogaster (fruit fly).		
OC	Eukaryota, Metazoa, Arthropoda, Tracheata, Hexapoda, Insecta,		
OC	Pterygota, Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherier S.E., Li P.W., Hoskins R.A., Galie R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.-R., Vandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.H., Blazel R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abail J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Baas A.U., Bakendell J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bernale B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mayes A.D., Dew I., Dietz S.M.,		
RA	Dodson K.C., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,		
RA	Flores C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,		
RA	Gleider A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman I.J., Hernandez J.R., Houck J.,		
RA	Hoselin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,		
RA	Jostali M., Kalush F., Karen G.H., Ke Z., Kemison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Krivitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Meitzi B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostretil A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nussekn D.R., Pachle J.M.,		
RA	Palazzo M., Peterson G.S., Pan S., Pollard J., Puti V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,		
RA	Spieler E., Spradling A.C., Stepleton M., Strong R., Sun E.,		
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,		
RA	Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster."		

```

RL   Science 287:2185-2195 (2000) .
DR   EMBL: AE003777; AA057135.1; -.
DR   FlyBase; FB00039834; CG3564.
DR   InterPro; IPR000092; PolyPrenyl_synth.
DR   Pfam; PF00348; PolyPrenyl_synth.1.
DR   PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR   PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
SQ   SEQUENCE 430 AA; 48759 MW; 3CF533CE66F3029 CRC64;

Query Match      33.8%; Score 742; DB 5; Length 430;
Best Local Similarity 37.2%; Pred. No. 4,2e-43;
Matches 167; Conservative 85; Mismatches 135; Indels 62; Gaps 6;

OY   3 SPALIRISIS-SRSIASLR-----SVTLRTATAPSLRLRCPETSPSSMAAASAGLV 57
DB   33 APTGTINHIETLESKSSVRLQPSATKESITYVGSPHKCSOSLOYSSKAMLRHSHSVH 92
OY   58 EPPDQPI-----INPLNVGPEMSNLTNSIRSLGSG--HPSLDTVAKYVQSEGKHIRP 111
DB   93 TQGPAGPREFQIDPYITLDDDKKPYDVAIVLLKSGISQSELDITIASYTDGCGKALRP 152
OY   112 LMLVLMQAOTEVAPRVQGEKVEVPEVNEGALPPEVLNDKNPDMNNRSGPLYDGEIEG 171
DB   153 MYTMLAKAIIN-----HLNN 168
OY   172 QTSNLSAQRRLAETEMTHIASLLHDVIDASERRRANPBGCAQFGKMAILAGDFLLG 231
DB   169 ESHQVHROQRIALFSEWVSHASLVHDVIDQSDFRRKGPSVNALMNNKRYTMAGDYLLS 228
OY   232 RASVALRLRNPEVIELATVIANVEGFMQKNVTVDALTEATQETPDYVLOKTYLK 291
DB   229 IASIMIALRSDVITVIVSQLITDLVQSEFMQLGSRFE-----NERFAHYLTKTIRK 281
OY   292 TASLIASCRASALLGGATPEVADAAYVAGNLLGALQIVDDMLDYVSATDLGKPGAD 351
DB   282 TASLIANLKKATAVIAQDDNVAAEVAFFQGNIGLAFQVDDMLDFVSTBQMGKPTAAD 341
OY   352 IQLGIATPALPAMGHAHELGPMTIKRKSDDGVDRARAEVLEYSGLKTKTALAEEYQK 411
DB   342 LKLGIATPVPVFPACKRYELNPMVRRRSESGDVERAEFLVHKSHGLEQTFLLAKHCNE 401
OY   412 ALDARFPESPAPKALBQLTDKVLTRSR 440
DB   402 AIRLAQELTESPYQKGLQVADLVINRMK 430

RESULT 3
O9FET8
ID   O9FET8      PRELIMINARY;          PRT;          416 AA.
AC   O9FET8;
DT   01-MAR-2001 (TREMBLrel. 16, Created)
DT   01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE   Geranyl diphosphat synthase.
GN   GPRS.
OS   Quercus robur (English oak).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eucosids I; Fagales; Fagaceae; Quercus.
OX   NCBI_TaxID=38942;
RN   [1]
RS   SEQUENCE FROM N.A.
RA   Zimmer W., Ochinski C.;
RT   "Identification of a prenyl transferase gene from Quercus robur.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AJ298245; CAC00852.1; -.
DR   InterPro; IPR000092; PolyPrenyl_synth.
DR   Pfam; PF00348; PolyPrenyl_synth.1.
DR   PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR   PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
SQ   SEQUENCE 416 AA; 46243 MW; 29142BD071680AF1 CRC64;

Query Match      33.5%; Score 734; DB 10; Length 416;

```

Best Local Similarity 41.8%, Pred. No. 1.4e-42;
Matches 161; Conservative 76; Mismatches 92; Indels 56; Gaps 6;

QY 67 NPLNLVGPENLNLSNIRSLGSHPSLDTWAKYYVQS--EGKHIRPLMWLLMAQATEVA 124
: : | : : : : | :
Db 77 DPFSIVADELSWANRLSMVVTEVPKLASAAEYFFFKGVGEGRFPPTVLLMATAWNIS 136
: : | : : : : | :
QY 125 ---PKVGWEKVVEPVNNEGLAPPEVLNDKNPDMNMRSGLTKDGETEGQTSNILASQR 181
: : | : : : : | :
Db 137 ILEPSLRG-----PGDAL-----TTTLRARQQ 158
: : | : : : : | :
QY 182 FLAETEMIHAASSLHDDVIDASETRRNAPSGNQAFGNKMALLAGDFLLGRASVALARLR 241
: : | : : : : | :
Db 159 RIAETEMIHVASLLHDDVDADRRRIGISLNFVMGNKLAVLAGDFLLSRACVALASLK 218
: : | : : : : | :
QY 242 NPVEIETLVIANLVEGEFMQLKNTVDDATEATQTFTDYLOKTYLKTASLIASKCR 301
: : | : : : : | :
Db 219 NTEVSVLLAKUVEHLVTGETMQMTTCQR-----SMEYMOKTYKTSLSINSCK 271
: : | : : : : | :
QY 302 ASALLGGATPEVADAAYAGRNLGAFOIVDDMLDYTVSATDLGKPAGADLQLGLATAPA 361
: : | : : : : | :
Db 272 AIALLGGOTSEVAMLAYEYGNLGLAYQLIDVDLFTGTSAASLGKSJSDIRHGIIITAPI 331
: : | : : : : | :
QY 362 LPAMKHAELGPMIKRKSPDGCDVERARELVEKSCLEKTKTALAEEYAQAOKALDAIRTPE 421
: : | : : : : | :
Db 332 LFAMEEFPOLREVVDVRGDDPANVDVALDYLGKSRGIQRARELAKKHANIAEAIDSUPE 391
: : | : : : : | :
QY 422 S-----PARKALEOLTDKVLTRSR 440
: : | : : : : | :
Db 392 SNDEEDVKSRRALDLTLTERTVIRTK 416
: : | : : : : | :

RESULT 4
Q9fT89 PRELIMINARY; PRT; 422 AA.

AC Q9fT89
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Geranyl diphosphate synthase.
GN GPS.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20521526; PubMed=11069698;
RP Bouvier F., Suire C., D'Harlingue A., Backhaus R., Camata B.;
RA "Molecular cloning of geranyl diphosphate synthase and
RT compartmentation of monoterpene synthesis in plant cells.";
RL Plant J. 24:241-252(2000).
DR EMBL; Y17376; CAC16849.1; .
DR InterPro; IPRO00092; Polyprenyl synt.
DR Pfam; PF00348; polyprenyl synt; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
SQ SEQUENCE 422 AA; 46402 MW; 5B332AA3F69C0588 CRC64;

Query Match 32.4%; Score 710.5; DB 10; Length 422;
Best Local Similarity 38.5%; Pred. No. 6.1e-41;
Matches 177; Conservative 84; Mismatches 132; Indels 67; Gaps 9;

QY 1 MASPALRTSI---SSRSIASLRSVTL-----RTASAPSLRLRCTPTRSPSSWAAAVS-- 51
: : | : : : : | :
Db 10 ISSFLNRNRFYSGSSQSLASHRAFIIPOQHGSCHSDSPHKGYVCRTTYSLKSPVFGFSHQ 69
: : | : : : : | :
QY 52 ---SASRLVEDPPNPQPLINPLNLVCPMSNLTNSRTSLLSGSHPSLDTVAKYVQS--EG 106
: : | : : : : | :
Db 70 LYHGSSSLVEEE-----LDPFSLVADELUSLNLKULREWMVLAEVPKLAASAEYFFKRGVQ 124
: : | : : : : | :
QY 107 KHIRPLMWLLMAQATEVAPKVGWEKVVEPVNNEGLAPPEVLNDKNPDMNMRSGLTKD 166

Db	210	GDFFLSRACGALAAALKONTEVALLATAVEHLVTGETMEITS-----STEORSMDYYM	262
Qy	286	QKTYUKATSLIAKSCRASALLGGATPEYADAAYVGNRLGLAFQIVDDMDLDYVSATDLG	345
Db	263	OKTYAKTASLISNSCKAAVAVLGTQTAEEVAVAFEPYGNRLGLAFQIIDLIDLPFGTSASLG	322
Qy	346	KPACADDLOGLATPALEFAMVHMHLELGMIRKRSDDQDVERAELEVKSDGLEKTRALA	405
Db	323	KGSLSDIHGHVITAIPLFAMEEPQQLREVVDVDEKDPNDVIALEYLGKSGIGORARELA	382
Qy	406	EYEAOKALDAIRTPPEP-----ARKKLQOLTEKVLTRSR	440
Db	383	MEHANLAAAAIGSLPETINEDYKSRRLIDLTRVITRKN	423

RESULT 6
Q9WU69
ID Q9WU69
PRELIMINARY;
PRT; 336 AA.

DT 01-NOV-1999 (TREMBlere1. 12, Created)
DT 01-NOV-1999 (TREMBlere1. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlere1. 21, Last annotation update)
DE Trans-prenyltransferase.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.
RA Rogic A., Geromet V., Chretien D., Parfait B., Kadhom N., Edery P.,
RA Leideau M., Rustin P., Munnich A., Appelqvist E.-L., Dallner G.,
RA Ernster L.?

RT encephalomyopathy.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases
 RN [2]
 RP SEQUENCE FROM N.A.

RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases
 DR EMBL, AF118855; AAAD24462.1; -.
 DR EMBL, BC026820; AAH26820.1; -.
 DR GGD, MG1:1889278; Tpf.

DR Pfam, PF00348; polyprenyl_synth, 1.
DR PROSITE, PS00723; POLYPRENYL_SYNTHET_1, 1.
DR PROSITE, PS00444; POLYPRENYL_SYNTHET_2, 1.
KW Transferase.

Query Match:	31.4%	Score 690;	DB 11;	Length 336;
Best Local Similarity:	39.8%	Pred. No. 1,1e-39;		
Matches 149;	Conservative 66;	Mismatches 107;	Indels 52;	Gaps 2

```
QY      67 NPLMLVGESEMNLSINISLSSGSPDLDTAKKYVYGEGHRLPWLVLMAQTAVAPK 126
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     15 DPKLGWDDLGLVEDIRKEHLISTRELKMSEYYFPDGKAFFPIIVLMARACNH-- 72
```

Db	73	-----HNNAREMQASGRSTALV	89	:: :: :
OY	187	TEMTIHAASLLHDVDVIDASETRRNAPSGNQAFGNKAILAGDFELIGRASVALARIINDEVI	246	

D5 90 AEMHTALVYDDVIDDASSRGRKHTYNNKIGGEKKAVALAGLLISASVALARIGNAVY 149

OY 247 ELLATVINLVGEFMQAKNTVDDAIENTAQTETFPDYLLQKTYIKTASLAKSGRASALL 306

QY 307 GCATPEVADAAAYAGRNGLAFOIYDDMLDYTSATDLGKPKAGADLQLGATAAPALFAWK 366
| : | : | : | : | : | : | : | : | : | : | : | : | :
jb 203 GCPDPVHEHIEIYYGKNVGIAFOLIDPVLDETSASDOMGKPTSDILKGATGVGFACQ 2822

Oy HHAELGPIIKKPSFDPCGVARELVKSQGLEKTRALAEVYNOKLDAIRTPESPARK 426
367 : |||::|||::|||::|||::|||:
Db QPFENNMMIMRFFSLPEDVDPRARYVLGSIDVGCTTLAAQQYCHKAVERIKLRPSTERD 322
263 : |||::|||::|||::|||::|||:

Oy ALLEQLTDKYLTRSR 440 .
427 : |||::|||::|||::|||:
Db ALLIQLSVESLTRDK 336
323 :

RESULT 7	
Q9Y2W5	
ID Q9Y2W5	PRELIMINARY;
Q9Y2W5	PRT; 376 AA.

DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE Trans-prenyltransferase.

OC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RA Røigård A, Gøromel V., Chretien D., Pafatit B., Kadhon N., Edsberg P.
RA Lehideau M., Rustin P., Munnich A., Appelkvist E.-L., Dallner G.,
RA Erster L.;
RT "Widespread coenzyme Q10 deficiency in familial mitochondrial

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases
 DR EMBL, AF118395; AAD28559.1; -
 DR InterPro: IPR000092; Polyprenyl_synt.
 DR Pfam: PF00348; polyprenyl_synt; 1.

DR PROSITE; PS00444; POLYBRENYL_SYNTHET_2; 1.
KM Transferrase.
SQ SEQUENCE 376 AA; 42080 MW; 91BBBE941812D193 CRC64;

Best Local Similarity 39.8%, Pred. No. 4e-39;
Matches 149, Conservative 65, Mismatches 108; Indels 52; Gaps 2;
67 NPLAVGPEMSNTSINISLLSGCHPSLDYIAKYVQSEGGHIRPLMLVLAQAQTEVAPK 126

```

Db      55 DPFKLGMRDLGLVEGRKPELSTTELKSESEYYPVKGAGAPPIVLVLMARCNH-- 112
Qy     127 VQGMKEKVEYEPVNEGGLAPPEVLNDKNPFMMNMRSGLPTKDEIEGQTSNIIASRLAEI 186
      ::|||::|

```

DQ 187 TEMHAASLHDDVIDASSETTRNAPSGNQAFGNMALLAGDFLIGRASVALARLRNEVI 246
| | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 130 AEMHTSLAVDDYIDDASSRGHGTNNKIKGEKKAVLAGDLLISAASILARIGNTTVI 189

```

Cy      247 ELMTAVIANVEGGEFMQKNTVDIAIEATITQEFSDYLYKTVIKTSLIAKSGRABALL 306
      : | | : | | | | : : : | | | | | | | | | | | : | : |
Db      190 SILNQVIEDLVRRGEFLQDGSKE-----NENFAYLEKTKTKTSLIAKSGKANSVL 242

```

Db 243 GCPDPVHEIAIYQYKXNGVIAFQIILDDVLDFTSCSDQMGKRTSADLKLGLATGVLFAQ 3020

Db 303 QPENNANIMMRFLSPGVNBARQYVLOSDBVQGTYYLAQQYCHEALREISKLRPSFERD 362

Qy 427 ALEQLETKVLTISR 440

RESULT
Q9FSW8

Oy 367 HHAELGPMTIKRPSFDPCGVSEAPARELYKSPGLETRALAEVQKALDAIRTPESPARK 426
 | || |||| || || || ||
Db 263 QPFENNAIMMRFSFLPGVDARQYVLQSDSVOQTTLAQQICAKAVREIRKLRFSTERD 322

Oy 427 ALEQLTKYLTRLRSR 440
 | || || || ||
Db 323 ALIQLSVESLTRLRK 336

RESULT 7	
Q9Y2W5	
ID Q9Y2W5	PRELIMINARY;
Q9Y2W5	PRT; 376 AA.

DT 01-NOV-1999 (TREMBlurel. 12, Created)
DT 01-NOV-1999 (TREMBlurel. 12, last sequence update)
DT 01-JUN-2001 (TREMBlurel. 17, last annotation update)
DE Trans-prenyltransferase.

OC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RA Røigård A, Gøromel V., Chretien D., Pafatit B., Kadhon N., Edsberg P.
RA Lehideau M., Rustin P., Munnich A., Appelkvist E.-L., Dallner G.,
RA Erster L.;
RT "Widespread coenzyme Q10 deficiency in familial mitochondrial

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases
 DR EMBL, AF118395; AAD28559.1; -
 DR InterPro: IPR000092; Polyprenyl_synt.
 DR Pfam: PF00348; polyprenyl_synt; 1.

DR PROSITE; PS00444; POLYBRENYL_SYNTHET_2; 1.
KM Transferrase.
SQ SEQUENCE 376 AA; 42080 MW; 91BBBE941812D193 CRC64;

Best Local Similarity 39.8%, Pred. No. 4e-39;
Matches 149, Conservative 65, Mismatches 108; Indels 52; Gaps 2;
67 NPLAVGPEMSNTSINISLLSGCHPSLDYIAKYVQSEGGHIRPLMLVLAQAQTEVAPK 126

```

Db      55 DPFKLGMRDLGLVEGRKPELSTTELKSESEYYPVKGAGAPPIVLVLMARCNH-- 112
Qy     127 VQGMKEKVEYEPVNEGGLAPPEVLNDKNPFMMNMRSGLPTKDEIEGQTSNIIASRLAEI 186
      ::|||::|

```

OY 187 TEMHAASLHDDVIDASETTRNAPSGNQAFGNMALLAGDFLIGRASVALARLRNEVI 246
||| ||| : | : : : ||| : : : ||| : : :
Db 130 AEMHTSALVHDDVIDDASSRGKHTNKIKWGEKKAVLAGDLIISAASILARGNTTVI 189

```

Cy      247 ELMTAVIANVEGGEFMQKNTVDIAIEATITQEFSDYLYKTVIKTSLIAKSGRABALL 306
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      190 SILNQVIEDLVRRGEFLQDGSKE-----NENFAYLEKTKFKTSLIAKSGKANSVL 242

```

Db 243 GCPDPVHEIAIYQYKXNGVIAFQIILDDVLDFTSCSDQMGKRTSADLKLGLATGVLFAQ 3020

Db 303 QPENNANIMMRFLSPGVNBARQYVLOSDBVQGTYYLAQQYCHEALREISKLRPSFERD 362

Qy 427 ALEQLETKVLTISR 440

RESULT
Q9FSW8

RA	Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA	Yasuda M., Tabata S.;
RT	"Complete genomic sequence of the filamentous nitrogen-fixing
RT	cyanobacterium Anabaena sp. strain PCC 7120.";
RL	DNA Res. 8:205-213(2001).
DR	EMBL; AP003581, BAB77620.1; --
DR	InterPro; IPR000092; Polyprenyl_synth.
DR	Pfam; PF00348; polyprenyl_synth; 1.
DR	PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR	PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW	Complete proteome.
SQ	SEQUENCE 323 AA; 35487 MW; BBE3BB0E1CA91094 CRC64;
Query Match 30.2%; Score 661.5; DB 16; Length 323;	
Best Local Similarity 38.4%; Pred. No. 9.7e-38;	
Matches 145; Conservative 72; Mismatches 102; Indels 59; Gaps	
Qy	61 PNQPLINPLNVLGPEMSNLTSNIRSLGSGHPSLDTVAKYYVQSGEKHTRPLMVLMAQA 120
Db	3 PATSLFTP---VEADLRILADNLKQLVGNRRHPILFAAAEHLFGAGGKRIRPAIVLLISRA 59
Qy	121 TEVAPKVGQEKVVEPVNIEGLAPPEVLNDKPNMNMRSGLTRKDGEIGQTSNILASQ 180
Db	60 T-----MLDQGITP-----RH 70
Qy	181 RLRAEITETIHAASLIHDDVIDASETRRNAPSGNQAFGNKMAILAGDFLLGRASVALARL 240
Db	71 RLRAEITETIHTASLHDDVDVDESEVRRGVTVHSLFGNRIAILAGDFLUFQAQSSWYLANL 130
Qy	241 RNFVEIETLATTIANLVEGFPMQLKNTVDDATEATQTFTFYDKTKYLTKTASLIAKSC 300
Db	131 DNLQVKKLSEVIMTDLATGSIQGLNRFDAI-----SIETYIEKSYKTKASLVANSS 183
Qy	301 RASALLGGATPEVADAAYAGRNIGLAFQIVDDMLDVTYSATDLGKPGACADLQGLATAP 360
Db	184 KAAGLLSEVSPETAELHYAGRHGHTAFQIVDDILOFTSTDTLGPVGSGLKSGNLTP 243
Qy	361 ALFAWKHSHSLGPMIKRKFSDPDEVRARELVEKSDGLEKTRALABEYQAQKALDAIRTFP 420
Db	244 VLPALAEKPYLEVLIEREFAQEGDLEQALELIQDSQIQSRELAHAHKTIAEHLATLP 303
Qy	421 ESPARKALEQLTKVLR 438
Db	304 PSESHQALIKIAEYAIR 321
RESULT 10	
P91093	PRELIMINARY; PRT; 393 AA.
AC	P91093;
DT	01-MAY-1997 (TrEMBLrel. 03, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Hypothetical 43.0 kDa protein.
OS	C24A11.9.
GN	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RC	MEDLINE=99069613; PubMed=9851916;
RA	None;
RT	"Genome sequence of the nematode C. elegans: a platform for
RT	investigating biology. The C. elegans Sequencing Consortium.";
RL	Science 282:2012-2018(1998).
RL	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RA	Maggi L.;
RT	"The sequence of C. elegans cosmid C24A11.";
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00443; AAB37678.2; -
DR InterPro; IPR000092; PolyPrenyl_syn.
DR Pfam; PF00348; PolyPrenyl_syn; 2.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 2.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 2.
KW Hypothetical protein.
SQ
SEQUENCE 393 AA; 43003 MW; 4C5PC2EAC07B5E2 CRC64;

Query Match 29.2%; Score 641.5; DB 5; Length 393;
Best Local Similarity 36.7%; Pred. No. 3.1e-36;
Matches 153; Conservative 74; Mismatches 125; Indels 65; Gaps 9;

QY 39 TSSRSSMAAASAS-----RLEPPDPNPPLNPLVGPENSLTSNI 83
DB 22 TSTSTSSDSSVASTAFVQEHVRQNDIMVQLIPQDESGAVENLADL-----NVTSN- 74
QY 84 RSLIGSGHPSLDYVAKYVYVSEGKIRPLMVLMAQATEVAPKVQGEKVEVVEGLA 143
DB 75 -----LGRMTHYVFOQGGKMLRPVSLMGNA CNSAARSISEVLIAMLSF--- 121
QY 144 PPEVLNDKPNPMNMNRSGPLTKDGEIGQTSNIIASQRLAEITEMIHAASLLHDDVIDA 203
DB 122 -----RSG-----IAHLS-VCONQYKIGMIEMIHSTASLVHDDVIDE 158
QY 204 SETRRNAPSGNQAFCNKAILAGDELIGRASVALARLNEVEVELLTVIANIVEGEFMQ 253
DB 159 ANTRRGASVVAAGNKSIVLVGDFILARATQILCSIGKPIVLSVMKSIIEVLGGEFMQ 218
QY 264 LKNTVDATETATQETFDYVLOQTYLKTASLLAKSCRASALLG-GATPEVADAAYVYGR 322
DB 219 MSTTPTATPV---DKMKAYIEKTHKRTASLPASSCRSAAILADSGDLKHLTAPEYGR 274
QY 323 NLGLAFOIVDMLDYTVSATDLGKPAQADLQLGLATAPALFAMKHAELGPMIKRFSQP 382
DB 275 NLGLAFOIADLDLDFIATDEMKGKPVADLKLGLATAPALVYACQYPELMTMLRKPKHD 334
QY 383 GDVERARELVESDGLKTRALAEVYQKALDARTPEEPAPKALEQLDXYLTS 439
DB 335 GDAEKAREIVNSDGMKTRRLIDSYQKAVEMASSLPN--RNESTEHLIKLAMSQS 369

RESULT 11
ID Q85Y08 PRELIMINARY; PRT; 245 AA.
AC Q85Y08;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, last annotation update)
DE REL8374P.
GN CG3684.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN-BERKELEY;
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuno J., Paclet J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY075463; AL68276.1; -
SQ
SEQUENCE 245 AA; 27384 MW; 66CCD7DD5A8A3A1D CRC64;

```

```

Query Match 28.5%; Score 625.5; DB 5; Length 245;
Best Local Similarity 49.6%; Pred. No. 1.9e-35;
Matches 125; Conservative 51; Mismatches 69; Indels 7; Gaps 1;

QY 189 MIHAASLLHDDVIDASSETRRNAPSGNQAFCNKMAILAGDFILIGASVALARLNEVEVEL 248
DB 1 MHSASLVHDDVIDSDRRGKPSVALMNHKKTMAQDYILSTASITANILRSDDVTIV 60
QY 249 LATVIANIVEGEFMQKNTVDATETATQETFDYVLOQTYLKTASLLAKSCRASALLG 308
DB 61 LQILTDLVQGEFMQLGSRETE-----NEPFAVYLTQYKRTASLIANLAKATAVIAQ 113
QY 309 ATPEVADAAYVYGRNLGLAFQIVDMLDYTVSATDLGKPAQADLQLGLATAPALFAMKHN 368
DB 114 ADDNVAEVAFOYGRNIGLAFQIVDMLDVFSSTRQMKRPTAADLKLGLATAPVLFACEKY 173
QY 369 AELGPMIKRFSQDGVRRARELVESDGLKTRALAEVYQKALDARTPEEPAPKAL 428
DB 174 PELNPMWRFRSEPDVRRARELVYKSHGLEOTRFLAKKCNEMAIRLAQELTESPYQKL 233
QY 429 EQLTDKVLTRSR 440
DB 234 QVADLVINRMK 245

RESULT 12
ID Q8S948 PRELIMINARY; PRT; 406 AA.
AC Q8S948;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, last annotation update)
DE Polyrenyl diphosphate synthase.
GN AT-TRANS-PT.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLOMBIA;
RA Hirooka K., Fukusaki E., Kobayashi A.;
RT "Polyrenyl Diphosphate Synthase Involved in Quinine Biosynthesis from
RT Higher Plant.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071514; BAB86941.1; -
SQ
SEQUENCE 406 AA; 44468 MW; 2FDD1877A569411 CRC64;

Query Match 26.7%; Score 585; DB 10; Length 406;
Best Local Similarity 33.6%; Pred. No. 2.6e-32;
Matches 135; Conservative 69; Mismatches 140; Indels 58; Gaps 5;

QY 35 RCTPTSRPSSMAAASASRL---VEPDPNP--LNPPLVGPENSLTSNIRSLIGS 89
DB 49 KAVPTSKSEIISLNGISQGVSPDLKQESKOPISLVTLFELVAVDQTLNDNLISVGA 108
QY 90 GHPSLDTVAKYVYVSEGKIRPLMVLMAQATEVAPKVQGEKVEVVEVNEGLAPPEVLN 149
DB 109 ENPVLISAABQIFGAGGRMRGVLVSHAT----- 140
QY 150 DKNPDMNMNRSGPLTKDGEIGQTSNIIASQRLAEITEMIHAASLLHDDVIDASSTRN 209
DB 141 -----AELAG-LKETTERHRLAEITEMIHSTASLIHDDVLDSDMRG 182
QY 210 ASGNQAFCNKAILAGDFILIGRASVALARLNEVEVELLTVIANIVEGEFMQLKTV 269
DB 183 KETVELEFGTRVAVLAGPFAQASWYLANLENLEVKLIQVLIKDFASGEIKQASSLFD 242
QY 270 DATETATQETFDYVLOQTYLKTASLLAKSCRASALLGATPEVADAAYVYGRNLGLAFQ 329
DB 243 -----CDTKLDEVLKSPYKTASLVASTKGAATFRRVPEPDTQGVYERGGKVLGSLFQ 295

```

RESULT 14					
Q9SHG4	PRELIMINARY;	PRT;	379 AA.		
ID	Q9SHG4				
AC	Q9SHG4;				
DT	01-MAY-2000 (TtEMBLrel. 13, Created)				
DT	01-MAY-2000 (TtEMBLrel. 13, Last sequence update)				
DT	01-DEC-2001 (TtEMBLrel. 19, Last annotation update)				
DE	Very similar to prenyl transferase.				
F20D23.25.					
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	euroside II; Brassicales; Brassicaceae; Arabidopsiis.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN=CV. COLUMBIA;				
RA	Federicipl N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,				
RA	Altafi H., Araujo R., Huizar R., Rowley D., Buehler E., Dunn P.,				
RA	Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,				
RA	Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,				
RA	Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;				
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.				
ENBL:	AC007651; AD50025.1; -				
DR	InterPro; IPRO00092; Polyprenyl_synth.				
DR	Fram; PF00348; polyprenyl synth; 1.				
DR	PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.				
DR	PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.				
KW	Transferase.				
SQ	SEQUENCE 379 AA; 41916 MW; AC07C09591SD9B9E CRC64;				
Query Match 26.2%; Score 574.5; DB 10; Length 379;					
Best Local Similarity 35.2%; Pred. No. 1.2e-31;					
Matches 127; Conservative 66; Mismatches 115; Indels 53; Gaps					
Qy	71 LVGEMSNLTNSRSLGSGHPSLDTVAKYVVOSEGRHIPLMVLMAQAATEVAPKVGW	130			
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :				
Dd	63 VVADDLQRINDLLSVGAENPVLISAEEQIFSAGGKRMPGLVFLVSRT-----	113			
Qy	131 EKKEVEVPWNEGLAPPELVNDKNPDMMNRSGPTKCGIEGTSTNILASORRLAEITEMI	199			
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :				
Dd	114 -----AELAG-LKELTVEHRRLGEEIEMI	136			
Qy	191 HAASLLHDVDIVSETERRNPASGNQAFGNKMAILAGDFLLGRASVALARLNPEVIELLA	250			
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :				
Dd	137 HTASLIHDDVDLDESDMRGRETVEHLFGTRVAVLAGDFMFAQASWYLANLENLEVIKLIS	196			
Qy	251 TVIANLVEGFPMQLNTVDDAIENATATQETFDYYLYKTLYKTASLIAKCSRASALLGCAT	310			
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :				
Dd	197 QVIKDFASGEIKQASSLFDCDVK-----LDDYMLKSYYKTASLVAASTKGAAIFSKVE	249			
Qy	311 PEVADAAYAVGRNLGLAFQIVDDMLDVTGSATDLGKPAGADLQLGLATAPALPAWKHAE	370			
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :				
Dd	250 SKVAQMYYQGKNGUGLSFQVVDDILDTFTQTTEQLGKPAANDLAGNTAPTAVIFALENEPR	309			
Qy	371 LGPMIKRFSPDPGDVERARELIVEKSDGLEKTRALAEYAQKALDAIRTFPESPARKALEQ	430			
Dd	: : : : : : : : : : : : : : : : : : : : : : : : :				
Dd	310 LREIISEFCPGSLEEAIEIVNRGGIKKAQELAKEKAELALKNKLNCLPRSGFRSALED	369			
Qy	431 L 431				
Dd	:				
Dd	370 M 370				
RESULT 15					
O64684	PRELIMINARY;	PRT;	297 AA.		
ID	O64684				
AC	O64684;				
DT	01-AUG-1998 (TtEMBLrel. 07, Created)				
DT	01-AUG-1998 (TtEMBLrel. 07, Last sequence update)				
DT	01-DEC-2001 (TtEMBLrel. 19, Last annotation update)				
DN	Putative trans-prenyltransferase.				
GN	AT2G34630.				

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Bento M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Peidlium T.V.,
 RA Buel C.R., Ketchum K.A., Lee J.-J., Ronning C.M., Koo H., Motilal K.S.,
 RA Cronin L.A., Sherr M., Vannken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carraena A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhagen G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.,
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004077; AAC26705.1; -
 DR InterPro; IPR000092; PolyPrenyl_synth.
 DR Pfam; PF00348; PolyPrenyl_synth. 1.
 DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 RN Transferrase
 QO SEQUENCE 297 AA; 32612 MW; 1669D01B3719F3FE CRG64;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 16, 2003, 06:38:46 ; Search time 2939 Seconds
(without alignments)
16368.472 Million cell updates/sec

Title: US-09-830-111A-1
Perfect score: 1653
Sequence: 1 ttttggggtcgaagtc.....atgcgaactcagaagtga 1653

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.btg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.un.*
28: em.vi.*
29: em.htg.hum.*
30: em.htg.inv.*
31: em.htg.other.*
32: em.htg.mus.*
33: em.htg.pln.*
34: em.htg.rod.*
35: em.htg.mam.*
36: em.htg.vit.*
37: em.sy.*
38: em.htgo.hum.*
39: em.htgo.mus.*
40: em.htgo.other.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1653	100.0	1653	6	E51323	E51323 Process for
2	242.8	14.7	2214	8	D84311	D84311 Schizosacch
3	231.4	14.0	2666	8	MC1496300	AJ496300 Mucor cir
4	222	13.4	1134	6	E13333	E13333 gDNA encodi
5	218.6	13.2	1492	8	D89265	D89265 Schizosacch
6	199.6	12.1	2097	8	SPBPU694	AL136498 S.pombe c
7	198.6	12.0	5371	12	CVU62637	U62637 Cloning vec
8	198	12.0	4176	6	A13038	A13038 Artificial
9	198	12.0	4176	12	SYNTRC99A	M27744 Cloning vec
10	198	12.0	4176	12	XXU13872	U13872 pTtc99a clo
11	198	12.0	4239	6	A29289	A29289 pSEC-Bpl co
12	197	11.9	4613	12	ASPBAD18	X81838 E.coli DNA
13	197	11.9	5075	12	U02444	U02444 Cloning vec
14	186.4	11.3	2799	3	AY075463	AY075463 Drosophil
15	186.2	11.3	4542	12	ASPBD24	X81837 E.coli DNA
16	186.2	11.3	6775	12	AY112733	AY112733 Broad hos
17	185.6	11.2	1262	10	AF118855	AF118855 Mus muscu
18	185.6	11.2	1552	10	BC026820	BC026820 Mus muscu
19	185.4	11.2	6124	12	AY112734	AY112734 Broad hos
20	184	11.1	4232	12	SYNLACIQ	L24193 Cloning vec
21	183	11.1	4735	12	SYNBETALAC	L24367 Cloning vec
22	182.8	11.1	1600	8	ORO298245	AJ298245 Quercus r
23	182.4	11.0	1402	8	ATH17376	Y17376 Arabidopsis
24	182.4	11.0	1408	8	AY093006	AY093006 Arabidops
25	180.8	10.9	966	6	AX043804	AX043804 Sequence
26	180.8	10.9	966	6	AX360800	AX360800 Sequence
27	180.8	10.9	966	6	AX428168	AX428168 Sequence
28	180	10.9	4991	12	EVPKK2333	X95387 Expression
29	179.8	10.9	1114	8	CSP243739	AJ243739 Citrus si
30	178	10.8	2681	12	SYNPUT18	D13250 Plasmid put
31	171.8	10.4	1485	9	AF118395	AF118395 Homo sapi
32	171.8	10.4	298750	1	AP005375	AP005375 Thermosyn
33	168.6	10.2	1641	9	AK024802	AK024802 Homo sapi
34	165	10.0	5241	6	AX027815	AX027815 Sequence
35	160.4	9.7	3302	6	A20357	A20357 plasmid pIG
36	160.4	9.7	3302	6	A20358	A20358 plasmid pIG
37	159.6	9.7	133859	1	D90899	D90899 Synchocyst
38	155.2	9.4	4009	6	A39734	A39734 Sequence 2
39	155.2	9.4	4009	6	AR069625	AR069625 Sequence
40	152.4	9.2	3423	6	AR9078	AR9078 Sequence 86
41	152.4	9.2	3423	6	AR195392	AR195392 Sequence
42	152.4	9.2	4633	6	A20355	A20355 plasmid pIG
43	152.4	9.2	4633	6	A20356	A20356 plasmid pIG
44	152	9.2	1097	6	AR008977	AR008977 Sequence
45	152	9.2	1097	6	AR075318	AR075318 Sequence

ALIGNMENTS

RESULT 1
E51323
LOCUS E51323 1653 bp DNA linear PAT 31-JAN-2002
DEFINITION Process for producing coenzyme Q10.
ACCESSION E51323
VERSION E51323.1 GI:18633579
KEYWORDS JP 2001061478-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1653)
AUTHORS Matsuda, H., Kawamukai, M., Vajima, A., Ikenaka, Y., Hasegawa, J. and Takahashi, S.
TITLE Process for producing coenzyme Q10
JOURNAL Patent: JP 2001061478-A 1 13-MAR-2001;

```
COMMENT
KANAKA CORP
OS Saisoella complicata
PN JP 2001061478-A/1
PD 13-MAR-2001
PF 24-AUG-1999 JP 1999237561
PR
PI HIDEYUKI MATSUDA, MAKOTO KAMAMUKAI, AKIYOSHI YAJIMA, PI
YASUHIRO IKENAKA,
PI JUNZO HASEGAWA, SATOMI TAKAHASHI
PC C12N15/09, C12N1/21, C12N9/12, C12N1/21, C12R1:19, C12N9/12, PC
C12R1:19)
PC C12N15/00
CC
CH
FT source
FT source
FEATURES
1. 1653
/organism="unidentified"
/db_xref="taxon:32644"
Location/Qualifiers
/organism="Saisoella complicata".
source
BASE COUNT 392 a 417 c 475 g 369 t
ORIGIN
Query Match 100.0%; Score 1653; DB 6; Length 1653;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTTTGGGGGCGAAAGCGGACGCGGTCAGAGTTCCGCTTGAGACGATAAAGGCTCG 60
DB 1 TTTTGGGGGCGAAAGCGGACGCGGTCAGAGTTCCGCTTGAGACGATAAAGGCTCG 60
QY 61 GAGATTGAGTCAGACCAAGCTTTGATCCGTAGGCTTACATCTTCAGCAATCATTTT 120
DB 61 GAGATTGAGTCAGACCAAGCTTTGATCCGTAGGCTTACATCTTCAGCAATCATTTT 120
QY 121 AATTCATATACCATGAGCTTACCATGAGCTTACCATGAGCTTACCATGAGCTT 180
DB 121 AATTCATATACCATGAGCTTACCATGAGCTTACCATGAGCTTACCATGAGCTT 180
QY 121 AATTCATATACCATGAGCTTACCATGAGCTTACCATGAGCTTACCATGAGCTT 180
DB 121 AATTCATATACCATGAGCTTACCATGAGCTTACCATGAGCTTACCATGAGCTT 180
QY 181 GCGCTCTGCGATCGGTTACCTTAAGAACAGCTCGGACCTTCACTAGCACTAAGATG 240
DB 181 GCGCTCTGCGATCGGTTACCTTAAGAACAGCTCGGACCTTCACTAGCACTAAGATG 240
QY 181 GCGCTCTGCGATCGGTTACCTTAAGAACAGCTCGGACCTTCACTAGCACTAAGATG 240
DB 181 GCGCTCTGCGATCGGTTACCTTAAGAACAGCTCGGACCTTCACTAGCACTAAGATG 240
QY 241 TACCCCGACGAGCCGCGCATTCGATTCATGAGCTGCTGCTGCTTCGCGCTGAGACT 300
DB 241 TACCCCGACGAGCCGCGCATTCGATTCATGAGCTGCTGCTGCTTCGCGCTGAGACT 300
QY 301 GGTGAGCTGATCCGATCAACCTCTCATCAATCCGCTCAACTTGTGCTCGAGAT 360
DB 301 GGTGAGCTGATCCGATCAACCTCTCATCAATCCGCTCAACTTGTGCTCGAGAT 360
QY 301 GGTGAGCTGATCCGATCAACCTCTCATCAATCCGCTCAACTTGTGCTCGAGAT 360
DB 301 GGTGAGCTGATCCGATCAACCTCTCATCAATCCGCTCAACTTGTGCTCGAGAT 360
QY 361 GTCGAATCTTAATCCAAATCCGATCTCTCCGCTGAGGACACCTTCTCTGACAC 420
DB 361 GTCGAATCTTAATCCAAATCCGATCTCTCCGCTGAGGACACCTTCTCTGACAC 420
QY 421 TGTCCCTAAATCTAATGTTCACTGAGGAGAAAGCATATTCGCTCGCTAGTACTGCT 480
DB 421 TGTCCCTAAATCTAATGTTCACTGAGGAGAAAGCATATTCGCTCGCTAGTACTGCT 480
QY 481 GATGCTCAGCGCGACGAGGATTGCGCAAAAGTTTCAGGCTTGGAGAGAGTCTGAGGT 540
DB 481 GATGCTCAGCGCGACGAGGATTGCGCAAAAGTTTCAGGCTTGGAGAGAGTCTGAGGT 540
QY 541 TCCGCTGAGACGAGGATCGACCAACAGAGGCTCATGACAAAGACCCGATATGAT 600
DB 541 TCCGCTGAGACGAGGATCGACCAACAGAGGCTCATGACAAAGACCCGATATGAT 600
QY 601 GAACATGAGTCAGACCATTAAGAGAGAGCGGATCGAGGACAGACGTCGAATAT 660
DB 601 GAACATGAGTCAGACCATTAAGAGAGAGCGGATCGAGGACAGACGTCGAATAT 660
QY 661 CCTGCTTCGCAACGCGGCTTGGCTGAGATCAAGAGATGATCATGACGATCACTCTT 720
DB 661 CCTGCTTCGCAACGCGGCTTGGCTGAGATCAAGAGATGATCATGACGATCACTCTT 720
```

```
DB 661 CCTGCTTCGCAACGCGGCTTGGCTGAGATCAAGAGATGATCATGACGATCACTCTT 720
QY 721 CCACGACGACGTTATCGACGCTTCGAGACCAAGAGAAACGACCATTCGAGAACGACG 780
DB 721 CCACGACGACGTTATCGACGCTTCGAGACCAAGAGAAACGACCATTCGAGAACGACG 780
QY 781 ATTGGAGAAACAAAGATGAGGATTTTGGCTGAGATTTCTTGTGGGACGCGGCTCTGTC 840
DB 781 ATTGGAGAAACAAAGATGAGGATTTTGGCTGAGATTTCTTGTGGGACGCGGCTCTGTC 840
QY 841 ATTGGAGAGTTGGCGCAATCCGAGAGGATGATGAGCTTTGGCTACTGTTATTCGAACTT 900
DB 841 ATTGGAGAGTTGGCGCAATCCGAGAGGATGATGAGCTTTGGCTACTGTTATTCGAACTT 900
QY 901 GGTGAGGAGAGTTTCATGACATTCGAAATTAATCTTTGATGATGATGATGATGATGATG 960
DB 901 GGTGAGGAGAGTTTCATGACATTCGAAATTAATCTTTGATGATGATGATGATGATGATG 960
QY 961 GACGAGAGAAACGTTTCATTCATATTCGAGAGACCTTACTGAAAGCTGCTCTGAT 1020
DB 961 GACGAGAGAAACGTTTCATTCATATTCGAGAGACCTTACTGAAAGCTGCTCTGAT 1020
QY 1021 TGCCAAAGTCGACAGAGAGAGGCGCTTCGAGTGTGCTACGCTGAGGTTGCTGATGC 1080
DB 1021 TGCCAAAGTCGACAGAGAGAGGCGCTTCGAGTGTGCTACGCTGAGGTTGCTGATGC 1080
QY 1081 TGCTTATCTTAACGAGAGAACTTTGTTGCAATTCAGATTCGTCAGACATGCTGCA 1140
DB 1081 TGCTTATCTTAACGAGAGAACTTTGTTGCAATTCAGATTCGTCAGACATGCTGCA 1140
QY 1141 CTACACCGCTTCGCTTACGACCTTCGATGACCGCGCGGTGAGACCTTCAGCTGCT 1200
DB 1141 CTACACCGCTTCGCTTACGACCTTCGATGACCGCGCGGTGAGACCTTCAGCTGCT 1200
QY 1201 GCGCACGCGCGCGGCTTCCTTGCATGAGAGACCAAGCGCGAGCTCGGTCCATGATCA 1260
DB 1201 GCGCACGCGCGCGGCTTCCTTGCATGAGAGACCAAGCGCGAGCTCGGTCCATGATCA 1260
QY 1261 GCGCAAGTTCTCTGACCCAGAGAGAGCTGACGAGCTGACGAGGTTGTCGAGAAAGTGA 1320
DB 1261 GCGCAAGTTCTCTGACCCAGAGAGAGCTGACGAGCTGACGAGGTTGTCGAGAAAGTGA 1320
QY 1321 TGGATTGAGAGAGAGAGAGGCTTCGAGAGAGATTCGAGAGAGGCTTGGATGAT 1380
DB 1321 TGGATTGAGAGAGAGAGAGGCTTCGAGAGAGATTCGAGAGAGGCTTGGATGAT 1380
QY 1381 TCGGAGCTTCCGAGAGATCCGCGACGAGAGGCTTGGAGAGATTCGAGAGAGGCTT 1440
DB 1381 TCGGAGCTTCCGAGAGATCCGCGACGAGAGGCTTGGAGAGATTCGAGAGAGGCTT 1440
QY 1441 GACTAGGTCAGAGATGAGATTCGAGCTCGTACCCGAGGATCTCTGATGATGACCTGCA 1500
DB 1441 GACTAGGTCAGAGATGAGATTCGAGCTCGTACCCGAGGATCTCTGATGATGACCTGCA 1500
QY 1501 GGCATGCAAGCTTGGCTTTTGGCGATGAGAGAAAGATTTTCAGCTCATGATTA 1560
DB 1501 GGCATGCAAGCTTGGCTTTTGGCGATGAGAGAAAGATTTTCAGCTCATGATTA 1560
QY 1561 ATCAAGACGAGAGAGAGGCTTGAATAAGAAATTTGCTGCGGACGATGAGCGGCTGT 1620
DB 1561 ATCAAGACGAGAGAGAGGCTTGAATAAGAAATTTGCTGCGGACGATGAGCGGCTGT 1620
QY 1621 CCACCTGACCCGATGCGGAACTCAGAGTGA 1653
DB 1621 CCACCTGACCCGATGCGGAACTCAGAGTGA 1653
RESULT 2
D84311 2214 bp DNA linear PLN 06-FEB-1999
LOCUS Schizosaccharomyces pombe gene for decaprenyl diphosphate synthase,
DEFINITION complete cds.
ACCESSION D84311
```

VERSION D84311.1 GI:1845554
 KEYWORDS dps; decaprenyl diphosphate synthase; ubiquinone biosynthesis.
 SOURCE Schizosaccharomyces pombe DNA.
 ORGANISM Schizosaccharomyces pombe
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;
 Schizosaccharomycetes.
 REFERENCE 1 (sites)
 AUTHORS Suzuki, K., Okada, K., Kamiya, Y., Zhu, X.F., Nakagawa, T., Kawamukai, M.
 and Matsuda, H.
 TITLE Analysis of the decaprenyl diphosphate synthase (dps) gene in
 fission yeast suggests a role of ubiquinone as an antioxidant
 JOURNAL J. Biochem. 121 (3), 496-505 (1997)
 MEDLINE 97279041
 REFERENCE 2 (bases 1 to 2214)
 AUTHORS Suzuki, K.
 TITLE Direct Submission
 JOURNAL Submitted (03-APR-1996) Kengo Suzuki, Shimane University, Faculty
 of Life and Environmental Science; Nishikawatsu 1060, Matsue,
 Shimane 690, Japan (E-mail:kawamuka@botan.shimane-u.ac.jp,
 Tel:0852-32-6587, Fax:0852-32-6587)
 COMMENT On Feb 21, 1997 this sequence version replaced gi:1304267.
 Sequence updated (19-Feb-1997) by:Kengo Suzuki.
 FEATURES
 Location/Qualifiers
 1. .2214
 /organism="Schizosaccharomyces pombe"
 /db_xref="taxon:4896"
 347. .1483
 /gene="dps"
 347. .1483
 /gene="dps"
 /codon_start=1
 /product="decaprenyl diphosphate synthase"
 /protein_id="BAA12314.1"
 /db_xref="GI:1845555"
 /translation="MTQYVYLKMRKLSLGVSRVSLVSTNRSNASHLKNLEQI
 SGRIOMNSSEFLIECKYTIAGCKMRPSVLMSKATSLCHGIDRSVVGDKYI
 DDDURFSTGQLPSQLAQIEMHLSLDDVIDHANVRGSPSSNVAFGNRR
 SILAGFILARASTAMARLNQVTELLATLADLRGFLQKNTMPSLSLEIKQSN
 FVYIEKSLFASDLISKSGASTILGQCSPTVATAGRGICGTAFOLMDVLDYT
 SKDDTLGKAAGADKLGLATAPVLPAWKYPGLGAMVNRNHPSPDIQRASLVECTD
 ALEQITWAKYIKAKDSLCLPDPSPARKALPALADKVIIRKK"
 BASE COUNT 676 a 408 c 403 g 727 t
 ORIGIN
 Query Match 14.7%; Score 242.8; DB 8; Length 2214;
 Best Local Similarity 57.4%; Pred. No. 6.6e-42;
 Matches 458; Conservative 0; Mismatches 337; Indels 3; Gaps 1;
 QY 649 GAGCTGAATATCTCGCTCGCAACGGGTTGGTGGATCACGGAGATGATCCATGC 708
 DB 679 GACGGGTCAAATCTCTCTCAATTGAGATTAGCACAAATAACCGAGATGATCCATAT 738
 QY 709 AGCATCACTCTCCAGCACAGCTATCGAGCTTCGAGACGACGAAACGACCATC 768
 DB 739 AGCAAGTTTCTGCTGATGACCATGTGATGATCAACGCTTAATGTCCGTAGAGCTCACCTTC 798
 QY 769 CGAAACACGACGATTCGGAACAAAGATGGCGATTTTCGGCTGGTGATTTCTTTGGGACG 828
 DB 799 AAGCAATGTTGCTTCGGTAAATCGACGGTCAATCTTCGGGTAATTTTCATCTTGACAG 858
 QY 829 GGCGTCTGTTGCAATGGCGAGTTGGCAATCCGAGAGTGATGAGCTTTTGGCTACTGT 888
 DB 859 GGCTTCGACTGCTATGCCCGCCCTTCGAAATCCCAAGTTACGGAGTTGTTAGTACAGT 918
 QY 889 TATTGCAACTTGGTTAGGGAGATTTCAGCTTGAATAAATCTGTTGATGATGCGAT 948
 DB 919 GATAGACAGCTTGGTTCCAGGTGAGTTTTCGAGCTAAAAAATACTATGATCCTTCATC 978
 QY 949 TGAGGCTACGGGACGAGGAAAGCTTCGATTACTATTTCGAGAGACTTACTTGAAGAC 1008
 DB 979 TTTGGAAA---TAAACAATCAAAATTTTGACTATTATATGAAAAAAGTTTTTGAAGAC 1035

QY 1009 TGGTCTCTGATTGCAAGTCGTGACGACCAAGTCGCTTCTGGTGGTCTAGCCTGA 1068
 DB 1036 AGCCAGTTTAAATTTCCAAAAGCTGCAAGGCTTCTACAATCTTCGACAAATGTTCTCTAC 1095
 QY 1069 GGTGCTGATGCTGCTTATGCTTACGGAAGAACCTTTGGTTTGGCAATTCAGATCGTCGA 1128
 DB 1096 TGAGCAACAGCTGCTGGAGAAATACGGTCGATGCTTGGTACTGCTTTTCAACTAATGA 1155
 QY 1129 CGACATGCTGACTACACCGTCTCCGCTACCGACCTCGGTAAGCCCGCGGTGAGACCT 1188
 DB 1156 TGACGTTGTTGGACTATACGTCGAAAGATGATCTTTAGGAAAGCGGCTGCTGCAGATTT 1215
 QY 1189 CAGCTCGGTCTCGCACCGCGGCGGCTCTTCGATGGAAGCACCACCGCGAGCTCGG 1248
 DB 1216 GAAGCTAGGGTTGGCTACAGCTCCCGTCTCTTTGCGATGGAAGAAAGTATCCAGAACTGG 1275
 QY 1249 TCCCATGATCAACGCGAAGTTCTCTGACCCAGGAGAGCTCGAGCGTGCAGCGAGTTGGT 1308
 DB 1276 TGCAATGATGTAAGATGATTCATCTCTGATATCAACGGGCTGTTCTTTGGT 1335
 QY 1309 CGAGAAAAGTATGATGATGTTGGAGAACGACGAGAGCTTGGCGGAGAGATGCCAGAAAGGC 1368
 DB 1336 TGAGTGCACTGATGCTATCGAGCAACCATCACTTGGGCAAAAAGATATATCAAAAAGC 1395
 QY 1369 GTTGGATGCAATTCGACGCTTCCGCGAGAGTCCGCGACGAGCTTTGGAGCAGTTGAC 1428
 DB 1396 CAAAGATTCCTTCTGTGCTCCCTGATTCACCTGCAAGGAGGCACTTTTTCGTTGGC 1455
 QY 1429 GGACAAAGTGTTCAGTAG 1446
 DB 1456 TGATAAGTAATAACGAG 1473
 RESULT 3
 LOCUS MCI496300 2666 bp DNA linear PLN 24-JUL-2002
 DEFINITION Mucor circinelloides f. lusitanicus isob gene for solanesyl
 pyrophosphate synthase, exons 1-3.
 ACCESSION AJ496300
 VERSION AJ496300.1 GI:21955861
 KEYWORDS isob gene; solanesyl pyrophosphate synthase.
 Mucor circinelloides f. lusitanicus.
 SOURCE Mucor circinelloides f. lusitanicus
 ORGANISM Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
 Mucor.
 REFERENCE 1
 AUTHORS Velayos, A.
 TITLE Carotenogenesis in Mucor circinelloides
 JOURNAL Thesis (2001) Department of Microbiologia y Genetica, Universidad
 de Salamanca, Salamanca, Spain
 REFERENCE 2
 AUTHORS Velayos, A., Fuentes, M., Aguilar, R., Eslava, A.P. and Iturriaga, E.A.
 TITLE Prenyl synthases in Mucor circinelloides
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 2666)
 AUTHORS Velayos, A.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUL-2002) Velayos A., Microbiologia y Genetica, Area
 Genetica, Facultad de Biologia, Universidad de Salamanca, Edificio
 Departamental, Avenida del Campo Charro s/n, Salamanca, E-37007,
 SPAIN
 FEATURES
 Location/Qualifiers
 1. .2666
 /organism="Mucor circinelloides f. lusitanicus"
 /strain="CBS277.49"
 /db_xref="taxon:29924"
 /note="forma: lusitanicus"
 515. .2091
 /gene="isob"
 join(515..1594,1650..1823,1893..2054)
 /gene="isob"
 /EC_number="2.5.1.11"
 /codon_start=1

```

/product="solanesyl pyrophosphate synthase"
/protein_id="CAD42868.1"
/db_xref="GI:21955862"
/translation="MFVVRQSKALSPVIRHRSIAASFHTTRAMLATPAKESDNDASAR
EKIHGNSFKSTVIGSPRLNITKRPVPTPTGSEANNQALSEQSLVKQSDRIID
PAKIVGDLKKEIKANISKLLSGHPLNTVARRHTSSGKVRPLVLLIQNTSIAD
KKSLSNIDYKSIDTPISHGLKNITTDVFRKLUHTYTSVTDGCTILPTORLAEIS
EMHITASLHDVDIDASMTRRNLPSANASFNKQAVLGDFLLAASLALRLNAEC
IELMATCIANLVEGEFMQLRNTEKESGKVKLSTFDHYMERKTYKTSLLAQSKAS
AVLGSTKEVANIAYDFGKNLGLAFQVLDMDLDFVTAAELGKPAQADLKLGLATAPV
LFAMEYRELEPLIKRKFVKDEKARDLVYSGDKKKTIDLAQIHCKLAPDALVYL
PASDARSALVOITKULTRRS"
exon
515..1594
/gene="150B"
/intron
1595..1649
/gene="150B"
/exon
1650..1823
/gene="150B"
/intron
1824..1892
/gene="150B"
/exon
1893..2054
/gene="150B"
/intron
2054..2059
/gene="150B"
/polyA_site
2091
/gene="150B"
BASE COUNT 741 a 692 c 538 g 695 t
ORIGIN
Query Match 14.0%; Score 231.4; DB 8; Length 2666;
Best Local Similarity 54.6%; Pred. No. 1.9e-39;
Matches 556; Conservative 0; Mismatches 396; Indels 67; Gaps 2;
QY 327 TCATCAATCCGCTCAACTGTGTCGCCGAGATGCAATCTTACATCAACATCCGAT 386
Db 813 TAAATGATCTCTCCCAACTAGTGGAGAGATCTCAAGAAATCAAGGCACATCTCCA 872
QY 387 CTCTCTCGGTTCAAGACACCTTCTCTGACACATGTCCTAAATCTATGTTCAAGTCTG 446
Db 873 AATTGCTGGGCGAGTGGACACCCGTTCTCAACACCGTGCGCCAGACACTTATAGCGGCG 932
QY 447 AGGGAAGCATTTTGTCCGCTCATGCTGCTGATGAGTCTCAGGCGAGGAGTTGCCG 506
Db 933 AGGTAAAGCATGTGCGACCATGCTGTGCTGTATAGCGGCGACACCGACATCGCAG 992
QY 507 CAAAAGTTCAGGGTTGGAGAGTCTGTGAGAGTTCCGGTGAACGAGGAGCTCGCACAC 566
Db 993 ACMAA-----AAGTGGCTGGCCAGCAACCTTGTACCAACGACGATTAAGCAG 1040
QY 567 CAGAGGTGCTCAATGACAAAGAACCCAGATATGATGAACATGAGTTCAGACCATTAAGCA 626
Db 1041 CTATATCGCATGAGCTCAAGAACATGACCAACGACCTGTTTGAACCGCAAGCTTACACT 1100
QY 627 AGGAGGGGAGATTCAGAGGACGAGCTGCAATATCTCGCTCGCAACGGCGGTTGGCTG 686
Db 1101 ACACACCCCTCCGCTGACAGATGAGGATGATCATCTTACCCACACGACGACGATTAAGCAG 1160
QY 687 AGATCAGGAGATGATCCATGAGCATCTCCTCCAGACGAGCTTATCGAGCTTCCG 746
Db 1161 AATATTCGAGATGATCTATAGCGGCTCATTTACTACACATGATGATGATGAGCGCTTCA 1220
QY 747 AGACCAAGAGAAACGACCATCCGAAACAGGCAATTCGAAACAAAGATGGCGATTTTG 806
Db 1221 TGACTCGGAGAACTTGCCGTCGCGCAATGCGTGGTTTGGCAACAAATGCGAGTCTGG 1280
QY 807 CTGGTGATTTTCTTTGGGACGGGCGTCTGTTGATTTGGCGAGTTGCCAATCCGAGAG 866
Db 1281 GCGGCGACTTCTTGGCGGCAAGGATCATTTGGCAAGACTCAGAAACGCGAGACT 1340

```

```

QY 867 TGATTGACCTTTGGCTACTGCTTATTTGCAAACTTGTTAGGAGAGTTGACAGTTGA 926
Db 1341 GCATGGAATTTGATGGCCACATGATATCGCAATCTGTGGAGGGCGAATTTATGCAATTGA 1400
QY 927 AAAATACCTTGATGATGATCCGATTTAGGCTTACGGCGACGCAAGAAACGTTGATTTACT 986
Db 1401 GAAACACCAAGAGAGGAATTCGGGCAAGGTGAAAACCTGAGCACCTTTGACATTAACA 1460
QY 987 TGCAGAACTTACTTAAAGACTGCGTCTTATTTGCCAAGTCGCGAGCAAGTGGCG 1046
Db 1461 TGGAGAAACCTTACATGAAGACGGGCACTGATTTGCCAAGCTGCGAGCGTGTGCGG 1520
QY 1047 TTTGGGTGTGTCTACGCTTGAAGTGTGCTGATGCTGTTATCTTACGAAAGAACTTG 1106
Db 1521 TACTGGGTGGCTCCACCAAGAAAGTCGCAACATTTGACATATGATTTTGGAAAGAACTTG 1580
QY 1107 GTTTGGCATTC----- 1118
Db 1581 GTTAGCCTTTACGGTAATTAACAAGCAATATCTTATTAACCCCTGTAAACACTTTTC 1640
QY 1119 -----AGATGCTGACGACATGCTCGACTACACCGTCTCGCTACCGACTGGTAA 1171
Db 1641 TCTGTATAGCTTTGTGATGATGATGCTGCTGATTTACAGATGACAGCAGAACTCGGCAAG 1700
QY 1172 CCGGCGGTGACAGACCTTCAGCTCGTCTGCGCACCGCGCCGCTCTTGCATGAG 1231
Db 1701 CCGGAGAGAGCAGATCTCAAGTGGGCTGGCCACAGCGCCGCTGTGTTGTTGCTTGGGAG 1760
QY 1232 CACCAAGCGGAGCTGGTCCATGATTCAGAGGCAAGTTCTTGACCCAGAGAGCTGCA 1290
Db 1761 GAGTATCCAGATTTGAGACCAATGATCAAGGAAAGTTCTGTAAAGGCGATGAGGA 1819

```

RESULT 4

```

E13333 LOCUS E13333 1134 bp DNA linear PAT 27-APR-1998
DEFINITION gDNA encoding prenyl diphosphate synthase.
ACCESSION E13333
VERSION E13333.1 GI:3252138
KEYWORDS JP 1997173076-A/1.
SOURCE Schizosaccharomyces pombe.
ORGANISM Schizosaccharomyces pombe
REFERENCE 1 (bases 1 to 1134)
AUTHORS Matsuda,H., Kawamuki,M. and Nakagawa,T.
TITLE PRODUCTION OF DIFFERENT KIND UBIOQUINONE
JOURNAL Patent: JP 1997173076-A 1 08-JUL-1997;
ALPHA-SHOKUHIN KK
COMMENT OS Schizosaccharomyces pombe
PN JP 1997173076-A/1
PD 08-JUL-1997
PF 27-DEC-1995 JP 1995351243
PI MATSUDA HIDEYUKI, KAWAMUKI MAKOTO, NAKAGAWA TSUYOSHI PC
C12N15/09, C12N9/10, C12P7/66, (C12N15/09, C12R1:19), (C12N9/09,
C12R1:645),
PC (C12N9/10, C12R1:865);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH key Location/Qualifiers
FT source 1..1134
FT 1..1134 /product='Prenyl diphosphate synthase'.

```

FEATURES

```

source 1..1134
location/Qualifiers
1..1134 /organism='Schizosaccharomyces pombe'
/db_xref='taxon:4896'

```


BASE COUNT	328 a	221 c	243 g	342 t
ORIGIN				
Query Match	13.4%; Score 222; DB 6; Length 1134;			
Best Local Similarity	56.6%; Pred. No. 2e-37;			
Mismatches	452; Conservative 0; Mismatches 340; Indels 6; Gaps 2;			
QY	649	GAGTGCATATCTCGCTCGCAACGGCGTGGCTGAGATCACGGAGATGATCCATGC	708	
DB	333	GACGGTCAAAATCTTCCTCTCAATGAGATTAGCACAAATAACCGAGATGATCCATAT	392	
QY	709	AGCATCACTCCTCCACGACAGCTTATCGACGCTTCGGAGACGAGAAACGACCATC	768	
DB	393	AGCAAGTTTCTGCATGACGATGATGATACGCTAATGTCCTGAGGCTCACCTTC	452	
QY	769	CGAAACACGAGCANTCGAAACAGATGCGAATTTTGGCTGGTGAATTTCTTGTGGACG	828	
DB	453	AAGCAATGTTGCTTTCGGTAAATCGACGGTCAATCCTTGGGGTAAATTCATCCTTCACG	512	
QY	829	GGCTCTGTGATTCGGAGGTTGGCAATCCGGAGGTGATGAGCTTTTGGCTACTGT	888	
DB	513	GGCTTCGACTGTATGGCCCGCTTCGAAATCCCAAGTTACGGAGTTGTTAGCTACAGT	572	
QY	889	TATTGAAACTTGGTTGAGGAGATTCATGCAGTTGAAATACTGTTGATGATCGAT	948	
DB	573	GATAGCAGACTTGGTTCGAGTGAGTTTTCGAGCTAAATAATACTATGATCCTTCATC	632	
QY	949	TGAGGCTACGGGACGAGAAACGTTTCGATTAATTTTCGAGAAGACTTACTTGAAGAC	1008	
DB	633	TTTGGAAA---TAAACAATCAAAATTTTGACTATTATATGAAAAAGTTTITGAA---	686	
QY	1009	TGGCTCCTTGATTCGCAAGTCGTGAGACGAAGTGGCTTCTGGGTGGTGCTACGCGTGA	1068	
DB	687	ACAGCAGATTATTTTCAAAAGCTGCAAGGCTTCTACAATCCTCGACAATGTTCTCTAC	746	
QY	1069	GGTTGCTGATGCTGTATGCTTACGGAAGAACCTTGGTTTGGCAATTCAGATCTCGA	1128	
DB	747	TGTAGCAACAGCTGCTGGGAATACGGTCGATGCTGTTGGTACTGCTTTTCAACATGGA	806	
QY	1129	CGACATGCTCGACTACACCGCTCTCCGCTACCGACCTCGGTAAAGCCCGGTCGACACCT	1188	
DB	807	TGACGTTGTTGACTATACGTCGAAAGATGATACTTTAGGAAAGCGCGTGTCCAGATT	866	
QY	1189	CMAGTCTGGTCTCGCACCGCGCGGCTCTTCGATGAAAGCACACGCGAGTCTGG	1248	
DB	867	GAACTAGGGTTGGCTACAGCTCCGCTCTTTCGATGGAAGAGTATCCAGAACTTGG	926	
QY	1249	TCCATGATCAAGCGCAAGTTCTGACCCAGGAGCGTTCGACGCTGACGCGAGTTGGT	1308	
DB	927	TGCAATGATTTGTAATAGATTCAATCATCTTCTGATATCCAACGGGCTGTTCTTTGGT	986	
QY	1309	CGAGAAAAGTGTGGATTGGAGAAGACGAGAGCTTTGGCGGAGGAGTATGCCAGAAAGC	1368	
DB	987	TGAGTCACATGCTATCGAGCAACCATCATCTTGGGCAAGAATATATCAAAAAGC	1046	
QY	1369	GTTGGATGCAATTCGACGTTCCGAGAGTCCGGACGGAAGCTTTGGAGCAGTTGAC	1428	
DB	1047	CAAAAGATTCCTTCTGTGCTCCCTGATTCACTGCAAGGAAGGCACCTTTTTCGGTTGC	1106	
QY	1429	GGACAAGTTGTTCACTAG	1446	
DB	1107	TGATAAGTAATAACGAG	1124	

RESULT 5
D89265
LOCUS D89265 1492 bp mRNA linear PLN 13-MAR-1998
DEFINITION Schizosaccharomyces pombe mRNA, partial cds, clone: SY 1754.
ACCESSION D89265
VERSION D89265.1 GI:1749737
KEYWORDS Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA,
SOURCE clone_lib:library of H. Nojima clone:SY1754.

ORGANISM	Schizosaccharomyces pombe Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.			
REFERENCE	1 (sites)			
AUTHORS	Yoshioka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.			
TITLE	Identification of open reading frames in Schizosaccharomyces pombe			
JOURNAL	CDNAB			
MEDLINE	DNA Res. 4 (6), 363-369 (1997)			
REFERENCE	2 (bases 1 to 1492)			
AUTHORS	Yoshioka,S.			
TITLE	Direct Submission			
JOURNAL	Submitted (15-NOV-1996) Sachiyo Yoshioka, Tsukita Cell Axis Project ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi, Shimokyo-Ku, Kyoto 600, Japan (E-mail: syoshi@cell.tsukita.jst.go.jp, Tel: +81-75-315-7913, Fax: +81-75-315-6420)			
FEATURES	Location/Qualifiers			
source	1..1492 /organism="Schizosaccharomyces pombe" /strain="PR745" /db_xref="taxon:4896" /clone="SY1754" /clone_lib="library of H. Nojima" 40..1110 /note="similar to Saccharomyces cerevisiae hexaphenyl pyrophosphate synthetase precursor, SWISS-PROT Accession Number p18900" /codon_start=1 /protein_id="BAA13926.1" /db_xref="GI:1749738"			
CDS	/note="similar to Saccharomyces cerevisiae hexaphenyl pyrophosphate synthetase precursor, SWISS-PROT Accession Number p18900" /codon_start=1 /protein_id="BAA13926.1" /db_xref="GI:1749738"			
BASE COUNT	453 a	282 c	291 g	466 t
ORIGIN				
Query Match	13.2%; Score 218.6; DB 8; Length 1492;			
Best Local Similarity	56.5%; Pred. No. 1.1e-36;			
Mismatches	450; Conservative 0; Mismatches 339; Indels 8; Gaps 2;			
QY	650	ACGTCGAATATCTCGCTCGCAACGGCGTGGCTGAGATCACGAGATGATCCATGCA	709	
DB	373	ACGGTCAAAATCTTCTCTTCAATGAGATTAGCAAAATACCGAGATGATCCATATA	432	
QY	710	GCATCACTCTCCACGACGAGTTATCGACGCTTCCGAGACGAGCAAAACGCCATCC	769	
DB	433	GCAAGTTTCTGCATGACGATGTTGATCAGCTAATGTCGTAGAGGCTCACCTTCA	492	
QY	770	GGAAACCGGCAATTCGAAACAAAGATGGCGATTTTGGCTGGTGAATTTCTTGTGGACGG	829	
DB	493	AGCAATGTTGCTTTCGGTAAATCGACGCTCAATCCTTTCGGGTAATTCATCTTGCACGG	552	
QY	830	GGCTCTGTTGCATTCGGAGGTTGCCAATCCGAGGTGATGAGCTTTTGGCTACTGTT	889	
DB	553	GCTTCGACTGTTATGGCCCGCTTCCGAAATCCCAAGTTCACGGAGTTGTTAGTACAGTG	612	
QY	890	ATTGCAAACTTGGTTGAGGAGAGTTTCATGCAAGTTCAAAAATACTGTTGATGATGCGATT	949	
DB	613	ATAGCAGACTTGGTTCGCGGTGAGTTTTCGAGCTAAAAATACTATGATCCTTCTACT	672	
QY	950	GAGGCTACGGCGACGAGAAACGTTTCGATTACTATTTCAGAAAGACTTACTTTGAAGACT	1009	
DB	673	TTGGAAA---TAAACAATCAAAATTTTGACTATTATATTGAAAAAAGTTTTTTTGAACA	729	
QY	1010	GGCTCTGTTGATTCGCAAGTCGTCGACGAGCAAGTCCGCTTCTGGGTGGTGTAGCCTGAG	1069	
DB	730	GCCAGTTTAAATTTCCAAAAAGCTGCAAGGCTTCTACAATCCTCGGACAATGTTCTCTCTACT	789	

QY 1070 GTTGCTGATGCTGCTTATGCTTACGGAAGAACCTTGGCATTCAGATGCTGCAC 1129
 Db 790 GTAGCAACAGCTGCTGGAATAACGCTCATGCTGCTGCTTTCACCTAAAGAT 849
 QY 1130 GACATGCTGACTTACACCGCTCCGCTACCGACCTGCTGAGCCCGCGTGCAGACTC 1189
 Db 850 GACGTTGTTGACTTATACGTCGAAAGATGATCTTATGAGAAAGCGCTGCTGCAATTTG 909
 QY 1190 CAGCTGCTGCTGCGACCGCGCGGCTCTTTCGATGAGAACACACCGCGAGCTCGT 1249
 Db 910 AAGCTAGGGTTGGCTACAGACTCCGCTCTTTCGATGAGAAAGATATCAGAACTGGT 969
 QY 1250 CCCATGATCAACCGGAGTCTTCTGACGAGAGAGCTGAGGCTGACGCGAGTTGGTTC 1309
 Db 970 GCAATGATGTAATAGATTAATCAATCTTCTGATATCAACGGGCTGCTTCTGCTT 1029
 QY 1310 GAGAAAGATGATGATGAGAAAGAGAGGCTTGGCGGAGAGATGCGCAGAAAGCG 1369
 Db 1030 GAGTCTCTGATGCTATCGAGAACCCATCATCTTGGCAATGAAATATATCAAAAA---- 1085
 QY 1370 TTGATGCAATTCGACGCTTCCCGGAGAGTCCGCGACGAGAGCTTTGGAGCATTTGACG 1429
 Db 1086 -AAGATTCCTCTGTGCTCTCCCTGATTCACCTCGACAGAGGCACTTTTGGCTTGGCT 1144
 QY 1430 GACAAAGTGTGACTAG 1446
 Db 1145 GATAAGTAAATACGAG 1161

RESULT 6
 SPBJ694 2097 bp DNA linear PLN 24-APR-2001
 LOCUS S.pombe chromosome II PCR product pJ694.
 DEFINITION AL136498
 ACCESSION AL136498.2 GI:13810218
 VERSION decaprenyl diphosphate synthase; dps.
 KEYWORDS fission yeast.
 SOURCE Schizosaccharomyces pombe
 ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;
 Schizosaccharomyces.
 1 (bases 1 to 2097)
 McDougall, R.C., Rajandream, M.A., Barrell, B.G., Saunders, D. and
 Harris, D.
 TITLE Direct Submission
 JOURNAL Submitted (17-JAN-2000) European Schizosaccharomycetes genome
 sequencing project, Sanger Centre, The Wellcome Trust Genome
 Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
 On Apr 26, 2001 this sequence version replaced gi:16714818.
 COMMENT Notes:
 Details of yeast sequencing at the Sanger Centre are available on
 the World Wide Web.
 (URL, http://www.sanger.ac.uk/projects/S_pombe/)
 During 1995 to 1996 about 66% of S. pombe chromosome 1 was
 sequenced by the Sanger Centre. The sequencing of the S. pombe
 genome is now being continued with funding from The European
 Commission. Fourteen European sequencing laboratories, including
 the Sanger Centre, are participating in the project.
 Protein coding regions (CDS) have been predicted with the help of
 computer analysis using the GeneFinder program in Pombase (an ACEDB
 database) with additional predictions for the branch-acceptor sites
 supplied by the program Splice. CAUTION: It is possible that for
 any individual CDS we may have underestimated or overestimated the
 number of introns/exons or we may not have chosen the correct
 splice donor/acceptor sites. CDS are numbered using the following
 system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2
 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the POSING database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions. PCR product pJ694
 is overlapped by cosmid c947, EMBL entry SPBC947, accession number
 AL021837, at the 5' end.

FEATURES
 source
 1..2097
 /organism="Schizosaccharomyces pombe"
 /strain="972h-"
 /db_xref="taxon:4896"
 /chromosome="II"
 /map="11L"
 1..100
 /note="nominal overlap with cosmid SPBC947, EM:AL021837 S.
 pombe chromosome 2"
 join(856..1830,1912..2073)
 /gene="dps"
 /note="SPBJ694.01; SPBJ4664.01"
 join(856..1830,1912..2073)
 /gene="SPBJ4664.01"
 /note="SPBJ694.01, len:378"
 /codon_start=1
 /label=dps
 /product="decaprenyl diphosphate synthase"
 /protein_id="CA86154.1"
 /db_xref="GI:6714819"
 /db_xref="SPRMBL:O43091"
 /translation="MIQVYVLKRMKLSLQKVRSTVLRSTTRNASHLKIHELQI
 SPGRIMLNSSEFLSECSXYITLQCKORPSVLVLSKATSLCHIDISVGDXYI
 DDDLRFSFTGQILPSQLRLQIEMHILSLDVIDYIDHNVGSPSSNVAGNRR
 SILAGNFIARASTAMARLRNPOTVELATVYIDLVNGEFLQKNTMPSLETKQSN
 PDVYIEKSLKTSATLISKSKASTILQCCSPYATAAGEGRICGTAFQMDVDVDT
 SKDPTLRKAGADIKLGLATAPVLFPAKKIPELGAMIVNPNHPSDIQRARSVECTD
 AIEQTIVAKREYIKKADSLCLPDSPARALFALADKVTTRKK"
 join(1030..1100,1179..1866)
 /gene="dps"
 /note="match to PF00348 polyprenyl_synth, Polyprenyl
 synthetases Score 137.19"
 1831..1911
 /gene="dps"
 /note="intron confirmed by EST"
 1831..1836
 /gene="dps"
 /note="gtaga, splice donor sequence"
 1896..1911
 /gene="dps"
 /note="ctacataatctag, splice branch and acceptor"
 BASE COUNT 682 a 355 c 383 g 677 t
 ORIGIN
 Query Match 12.1%; Score 199.6; DB 8; Length 2097;
 Best Local Similarity 57.6%; Pred. No. 1.3e-32;
 Matches 377; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
 QY 649 GACGTGCAATATCCCGCTCGCAACGCGGTGGCTGATGATCGAGATGATCCAGC 708
 Db 1188 GACGGGTCAAAATCTTCTTCAATTAGATGACAAATATACCGAGATGATCAAT 1247
 QY 709 AGCATCACTCTCCACGACGAGATGATGACGCTCCGAGACGACAAACGACCAATC 768
 Db 1248 AGCAAGTTGCTGCAATGACATGATGATGATGATGATGATGATGATGATGATG 1307
 QY 769 CGAATCCAGGATTCGAAACAGAGTGGCATTTTGGCTGATGATTTCTTGTGGAGC 828
 Db 1308 AAGCAATGTTGGCTTGGTAAATGACGAGTCAATCTTCCGGGTAAATTTCAATCCTT 1367
 QY 829 GGGCTGTGATGATGAGAGTTGCGCAATCCGAGATGATGATGATGATGATGATGAT 888
 Db 1368 GGGCTGACTGTATGAGCGCGCTTCGAAATCCCAAGTTACGAGTTGATGATGATG 1427
 QY 889 TATTCGAACTGGTTGAGGAGATGATGATGATGATGATGATGATGATGATGATGAT 948
 Db 1428 GATGACAGCTTTGTTCCAGGATGATGATTTTTCAGCTATAAAAATGATGATGATC 1487

```
QY 949 TGAGGCTACGGCAGCAGGAAGCTTCGATTACTATTTCGAGAAGACTTACTTGAAGAC 1008
Db 1488 TTTGGAAA---TAAACAACTCAAAATTTTGACTATTATTAATGAAAAAGTTTTTTGAAAAAC 1544
QY 1009 TCGCTCCTTGATGCCAAGTCGTGCAGAGAAGTGCCTTCTCGGTGGTGTCTACGCCCTGA 1068
Db 1545 AGCCAGTTTAATTTCCAAAAGCTGCAAGGCTTCTACAATCCTCGGACAATGTTCTCTAC 1604
QY 1069 GGTTCGTGATGCTTATGCTTACGGAAGAACCTTGGTTGGCATTCAGATTCGTCGA 1128
Db 1605 TGTAGCAACAGCTGCTGGAGAATACGGTCGATCGATTGGTACTGCTTTTCAACTAATGGA 1664
QY 1129 CGACATGCTCGACTACACCGTCTCCGCTACCGACTCGGTAAAGCCCGCGTGCAGACCT 1188
Db 1665 TGACGTGTTGACTATACGTCGAAAGATGATCTTTAGGAAAGCGGCTGGTGAGATTT 1724
QY 1189 CCAGCTCGGTCTCGCCACCGCGCGGCTCTTCGATGGAAGCACACCGCGAGCTCGG 1248
Db 1725 GAAGCTAGGTTGGCTACAGCTCCCGTCTCTTTGCGATGGAAGAGTATCCAGACTTGG 1784
QY 1249 TCCATGATCAAGCGGCAAGTTCTCTGACCCAGGAGAGCTCGAGCGTGCACGCGA 1302
Db 1785 TGCATGATTGTGAATAGATTCAATCATCTCTCTGATATCAACGCGGTAAGAGA 1838
```

```
RESULT 7
CVU62637
LOCUS Cloning vector pBAD-GFPuv, 5371 bp DNA linear SYN 14-AUG-1996
DEFINITION Cloning vector pBAD-GFPuv, complete sequence.
ACCESSION U62637
VERSION U62637.1 GI:1490531
KEYWORDS
SOURCE Cloning vector pBAD-GFPuv.
ORGANISM Cloning vector pBAD-GFPuv.
REFERENCE 1 (bases 1 to 5371)
AUTHORS Crameri,A., Whitehorn,E.A., Tate,E. and Stemmer,W.P.
TITLE Improved green fluorescent protein by molecular evolution using DNA shuffling
JOURNAL Nat. Biotechnol. 14 (3), 315-319 (1996)
MEDLINE 98294348
PUBMED 9630892
REFERENCE 2 (bases 1 to 5371)
AUTHORS Crameri,A. and Kitts,P.A.
TITLE pBAD-GFPuv complete sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5371)
AUTHORS Kitts,P.A.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1996) CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA
COMMENT This vector can be obtained from CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA. To place an order call (415) 424-8222 or (800) 662-2566, extension 1. International customers, please contact your local distributor. For technical information, call (415) 424-8222 or (800) 662-2566, extension 3.
This sequence has been compiled from information in the sequence databases, published literature and other sources, together with partial sequences obtained by CLONTECH. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES
source Location/Qualifiers
1..5371
/organism="Cloning vector pBAD-GFPuv"
/db_xref="taxon:50707"
complement(96..974)
/gene="araC"
complement(96..974)
/gene="araC"
/note="PID: g455167"
/codon_start=1
```

```
/transl_table=11
/product="araC protein"
/protein_id="AAC53662.1"
/db_xref="GI:1490532"
/translation="MAEAQNDPLPGYSFNAHLVAGLTPIEANGYLDFFIDRLPLMKMG
YLNLTIROGVVKNQREFVCRPGDILLFPGEIHVYGRHPEAREMYHQWVYFRPRA
YHWEHLNWPISIFANTGFFRPDEAHQPHFSDLFGQIINAGQEGREYSELAINLLEQLL
LRKMAINESLHPPMDNRVREACQIISDHLASNFDIASVAQHVCLSPSKRLSHLFROQ
LGISVLSWREDQRI1SOAKLLLTSTRPIATVGRNVGFDQDLYFSRVFKCTGTGASPSF
RAGCEKVNDAVKLS"
gene 1342..2061
CDS 1342..2061
/feature="gfpuv"
/note="gfpuv is the GFP variant called 'cycle 3'; Allele:
AC2; green fluorescent protein variant"
/transl_table=11
/codon_start=1
/product="GFPuv"
/protein_id="AAC53663.1"
/db_xref="GI:1490533"
/translation="MASGEBELFTGVVPIVLVDGDNVGHKFSVSGEGEDATYGLKT
LKFICTTGKLPVMPVPTLVTTFYGVQCFRSYPDHMKRHDFFKSAPEGVYOEFTISFK
DDGNVKTAEVKEFGDTLVNRIELGIDFKEDGNILGHKLEYNVNSHNVYITADKQK
GIKANFKRHNIEDSGVQLADHYQQNTPIGDGFPVLLPDNHYLSTQSALSQDPNEKRDH
MVLLEFVTAAGTTHGMDELYK"
gene 2636..3496
CDS 2636..3496
/feature="bla"
/function="confers resistance to ampicillin"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AAC53664.1"
/db_xref="GI:1490534"
/translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
IELDLSNGKILEDFRPEEFPMWSTFKLLCGAVLSRVDAQDQGLRRIRHYSQNDLVE
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTIGPKELTAFLHNMGGDHVTRL
DRWEPELNEAIPIQNDKDTTPMAAMATTLKLTGELLTLASRQQLTDMWEADKVAQPL
LRSALPAGWFIADKSGAGERSGIIAALGPDGKPSRIVVIYTTGSAQTMDERNRQIA
EIGASLLIKHW"
BASE COUNT 1369 a 1368 c 1300 g 1334 t
ORIGIN
Query Match 12.0%; Score 198.6; DB 12; Length 5371;
Best Local Similarity 98.0%; Pred. No. 2.2e-32;
Matches 201; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1449 CAAGATAGGAATTCGAGTCGGTACCCGGGATCCTCTAGAGTCGACCTCGAGGCATGCA 1508
Db 2055 CAAATAATGAATTCGAGTCGGTACCCGGGATCCTCTAGAGTCGACCTCGAGGCATGCA 2114
QY 1509 AGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCCTGATACAGATTAAATCAGAAC 1568
Db 2115 AGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCCTGATACAGATTAAATCAGAAC 2174
QY 1569 GCAGAAGCGTCTGATAAACAAGATTTTCCTCGCGGCAGTACGCGGTGCTCCCACTG 1628
Db 2175 GCAGAAGCGTCTGATAAACAAGATTTTCCTCGCGGCAGTACGCGGTGCTCCCACTG 2234
QY 1629 ACCCATGCGGAACTCAGAAAGTGA 1653
Db 2235 ACCCATGCGGAACTCAGAAAGTGA 2259
RESULT 8
LOCUS Al3038
DEFINITION Artificial sequence of plasmid pTtc99A.
ACCESSION Al3038
VERSION Al3038.1 GI:491502
KEYWORDS
```

SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE artificial sequences.
AUTHORS 1 (bases 1 to 4176)
TITLE Amann, E. and Abel, K.J.
JOURNAL Expression vectors for the production of non-fusion proteins in
BEHRINGWERKE Aktiengesellschaft
FEATURES
SOURCE 1. .4176
/db_xref="taxon:32630"
BASE COUNT 1009 a 1059 c 1118 g 990 t
ORIGIN
Query Match 12.0%; Score 198; DB 6; Length 4176;
Best Local Similarity 100.0%; Pred. No. 3e-32;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1456 GGAATTCAGCTCGGTACCGCGGGATCCTCTAGAGTCGACCTGCAAGCATGCAAGCTTGG 1515
Db 269 GGAATTCAGCTCGGTACCGCGGGATCCTCTAGAGTCGACCTGCAAGCATGCAAGCTTGG 328
QY 1516 CTGTTTGGCGGATGAGAGAAATTTTTCAGCTGATACGATTAATATGAAACGCAAG 1575
Db 329 CTGTTTGGCGGATGAGAGAAATTTTTCAGCTGATACGATTAATATGAAACGCAAG 388
QY 1576 CGGTCTGATTAAGAAAGATTTTCTGCGCGGAGTAGCGGAGTGTCCACCTGACCCCAT 1635
Db 389 CGGTCTGATTAAGAAAGATTTTCTGCGCGGAGTAGCGGAGTGTCCACCTGACCCCAT 448
QY 1636 GCCGAACCTCAGAAAGTGAA 1653
Db 449 GCCGAACCTCAGAAAGTGAA 466
RESULT 9
LOCUS SYNRC99A 4176 bp DNA linear SYN 29-JUN-1995
DEFINITION Cloning vector pTRC99A, complete sequence.
ACCESSION M22744
VERSION M22744.1 GI:340740
KEYWORDS cloning vector.
SOURCE Cloning vector pTRC99A.
ORGANISM Cloning vector pTRC99A.
REFERENCE 1 (sites)
AUTHORS Amann, E. and Brosius, J.
TITLE 'ATG vectors' for regulated high-level expression of cloned genes
JOURNAL in Escherichia coli
MEDLINE Gene 40 (2-3), 183-190 (1985)
PUBMED 3007288
REFERENCE 2 (bases 1 to 4176)
AUTHORS Amann, E., Ochs, B. and Abel, K.J.
TITLE tightly regulated tac promoter vectors useful for the expression of
unfused and fused proteins in Escherichia coli
JOURNAL Gene 69 (2), 301-315 (1988)
MEDLINE 89172074
PUBMED 3069586
FEATURES
SOURCE Location/Qualifiers
1. .4176
/organism="Cloning vector pTRC99A"
/db_xref="taxon:40992"
/lab_host="Escherichia coli"
/note="derived from pKK233-2"
1. .17
/note="derived from cloning vector pBR322"
misc_feature 18. .263
promoter /note="tac promoter from pKK233-2"
/citation=[1]
misc_feature 264. .270

misc_feature /note="NcoI/EcoRI linker"
271. .321
/note="derived from cloning vector pUC18"
misc_feature 329. .753
/note="5S RNA, T1, T2, rRNA"
misc_feature 754. .2932
/note="derived from cloning vector pBR322"
misc_feature 2933. .2940
/note="BgIII linker"
misc_feature 2933. .2938
/note="BglII linker"
misc_feature 2939. .2945
/note="EcoRI linker"
misc_feature 2946. .2976
/note="derived from plasmid RP4"
misc_feature 2977. .4163
/note="lacI-q region"
misc_feature 4164. .4170
/note="EcoRI linker"
misc_feature 4171. .4176
/note="BglIII linker"
BASE COUNT 1009 a 1059 c 1118 g 990 t
ORIGIN
Query Match 12.0%; Score 198; DB 12; Length 4176;
Best Local Similarity 100.0%; Pred. No. 3e-32;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1456 GGAATTCAGCTCGGTACCGCGGGATCCTCTAGAGTCGACCTGCAAGCATGCAAGCTTGG 1515
Db 269 GGAATTCAGCTCGGTACCGCGGGATCCTCTAGAGTCGACCTGCAAGCATGCAAGCTTGG 328
QY 1516 CTGTTTGGCGGATGAGAGAAATTTTTCAGCTGATACGATTAATATGAAACGCAAG 1575
Db 329 CTGTTTGGCGGATGAGAGAAATTTTTCAGCTGATACGATTAATATGAAACGCAAG 388
QY 1576 CGGTCTGATTAAGAAAGATTTTCTGCGCGGAGTAGCGGAGTGTCCACCTGACCCCAT 1635
Db 389 CGGTCTGATTAAGAAAGATTTTCTGCGCGGAGTAGCGGAGTGTCCACCTGACCCCAT 448
QY 1636 GCCGAACCTCAGAAAGTGAA 1653
Db 449 GCCGAACCTCAGAAAGTGAA 466
RESULT 10
LOCUS XXU13872 4176 bp DNA circular SYN 07-DEC-1994
DEFINITION pTRC99A cloning vector, complete sequence.
ACCESSION U13872
VERSION U13872.1 GI:595782
KEYWORDS 5S; beta-lactamase; lac repressor protein.
SOURCE unidentified cloning vector.
ORGANISM unidentified cloning vector.
REFERENCE 1 (bases 1 to 4176)
AUTHORS Malone, J.A.
TITLE pTRC99A: A cloning vector for inducible expression of cloned
inserts using the trc promoter
JOURNAL unpublished (1994)
REFERENCE 2 (bases 1 to 4176)
AUTHORS Amann, E., Ochs, B. and Abel, K.J.
TITLE tightly regulated tac promoter vectors useful for the expression of
unfused and fused proteins in Escherichia coli
JOURNAL Gene 69 (2), 301-315 (1988)
MEDLINE 89172074
PUBMED 3069586
REFERENCE 3 (bases 193 to 222)
AUTHORS Brosius, J., Bille, M. and Storella, J.
TITLE Spacing of the -10 and -35 regions in the tac promoter. Effect on
its in vivo activity
JOURNAL J. Biol. Chem. 260 (6), 3539-3541 (1985)
MEDLINE 85131154

Db 512 GCCAACTCAGAGTGAA 529

RESULT 12

LOCUS ASPB18 4613 bp DNA circular SYN 25-OCT-1995

DEFINITION E.coli DNA for pBAD18 cloning vector.

ACCESSION X81838

VERSION X81838.1 GI:551423

KEYWORDS arac gene; bla gene; expression vector.

SOURCE synthetic construct.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 4613)

AUTHORS Guzman, L.M.

TITLE Direct Submission

JOURNAL Submitted (22-SEP-1994) L.M. Guzman, Harvard Medical School, Dept of Microbiology & Molecular Genetic., 200 Longwood Ave., Boston MA 02115, USA

REFERENCE 2 (bases 1 to 4613)

AUTHORS Guzman, L.M., Bellin, D., Carson, M.J. and Beckwith, J.

TITLE Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter

JOURNAL J. Bacteriol. 177 (14), 4121-4130 (1995)

MEDLINE 95332220

PUBMED 7608087

COMMENT Related sequences: J01641 and M10196 (bases 1-1300); M77749 (bases 1357-2296; 4537-4613); SYNPR322 (bases 2297-2780); Gene 27, 183, 1984 (bases 2780-3238).

FEATURES

source 1..4613

Location/Qualifiers

/organism="synthetic construct"

/db_xref="taxon:32630"

/lab_host="Escherichia coli"

/plasmid="pBAD24"

gene 96..974

/gene="araC"

mRNA 96..974

/gene="araC"

misc_feature 1003..1020

/note="operator O2"

promoter 1125..1153

/note="Pc promoter"

misc_feature 1161..1182

/note="operator O1"

misc_feature 1204..1217

/note="CAP site"

misc_feature 1213..1251

/note="operator I2 + I1"

promoter 1250..1277

/note="PBAD promoter"

misc_feature 1300..1362

/note="multiple cloning site"

gene 1363..1788

/gene="trnB"

terminator 1363..1788

/gene="trnB"

promoter 1843..1849

/note="bla pJ promoter"

gene 1881..2744

/gene="bla"

mRNA 1881..2744

/gene="bla"

rep_origin 2780..3238

/note="ori3 origin"

rep_origin 3244..3941

/note="pBR322 origin"

BASE COUNT 1124 a 1214 c 1150 g 1125 t

ORIGIN

Query Match 11.9%; Score 197; DB 12; Length 4613;

Best Local Similarity 100.0%; Pred. No. 4.9e-32;

Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1457 GAATTCGAGCTGGTACCCGGGGATCTCTAGAGTGCATTCGACGATGCAAGCTTGGC 1516

DB 1306 GAATTCGAGCTGGTACCCGGGGATCTCTAGAGTGCATTCGACGATGCAAGCTTGGC 1365

QY 1517 TGTTTTGGGGATGAGAGAATTTTTCAGCTCGATACAGATTAAATCGAAGCGGAAGC 1576

DB 1366 TGTTTTGGGGATGAGAGAATTTTTCAGCTCGATACAGATTAAATCGAAGCGGAAGC 1425

QY 1577 GGTCTGATATAACAGAAATTTGCTGGCGCAGTACGCGGTGTCTCCACCTGACCCCATG 1636

DB 1426 GGTCTGATATAACAGAAATTTGCTGGCGCAGTACGCGGTGTCTCCACCTGACCCCATG 1485

QY 1637 CCGAACTCAGAGTGAA 1653

DB 1486 CCGAACTCAGAGTGAA 1502

RESULT 13

LOCUS U02444 5075 bp DNA circular SYN 29-MAR-1996

DEFINITION Cloning vector PK388-1, complete sequence.

ACCESSION U02444

VERSION U02444.1 GI:413810

KEYWORDS

SOURCE Cloning vector PK388-1.

ORGANISM Cloning vector PK388-1.

REFERENCE 1 (bases 1 to 5075)

AUTHORS Kites, P.A.

TITLE Unpublished

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 5075)

AUTHORS Brosius, J.

TITLE Expression vectors employing lambda-, trp-, lac- and lpp-derived promoters

JOURNAL (in) Rodriguez, R.L. and Denhardt, D.T. (Eds.); VECTORS: 205-225; (1988)

REFERENCE 3 (bases 1 to 5075)

AUTHORS Kites, P.A.

TITLE Direct Submission

JOURNAL Submitted (07-OCT-1993) Paul A. Kites, CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303, USA

COMMENT This vector can be obtained from CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303, USA. To place an order call (415) 424-8222 or (800) 662-2566, extension 1. International customers, please contact your local distributor. For technical information, call (415) 424-8222 or (800) 662-2566, extension 3. This sequence was compiled by Jurgen Brosius. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.

FEATURES

source 1..5075

Location/Qualifiers

/organism="Cloning vector PK388-1"

/db_xref="taxon:31816"

BASE COUNT 1210 a 1311 c 1418 g 1136 t

ORIGIN

Query Match 11.9%; Score 197; DB 12; Length 5075;

Best Local Similarity 91.3%; Pred. No. 4.9e-32;

Matches 209; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1425 TGAAGCAAGGTGTGACTAGTCAAGTATGAGATTGGATCCGGGGATCTT 1484

DB 309 TTATCATATAGAGGTTTAAACATGCGAGCTGAATTCGAGTCCGATCCCGGGATCTT 368

QY 1485 CTAGAGTCAGCTGCAGGATGCAAGCTTGGCTTTTGGCGGATGAGAGAATTTTCA 1544

DB 369 CTAGAGTCAGCTGCAGGATGCAAGCTTGGCTTTTGGCGGATGAGAGAATTTTCA 428

TITLE Direct Submission
JOURNAL Submitted (22-SEP-1994) L.M. Guzman, Harvard Medical School, Dept of Microbiology & Molecular Genet., 200 Longwood Ave., Boston MA 02115, USA

REFERENCE 2 (bases 1 to 4542)
AUTHORS Guzman, L.M., Belin D., Carson, M.J. and Beckwith, J.
TITLE Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter
J. Bacteriol. 177 (14), 4121-4130 (1995)

JOURNAL MEDLINE 95332220
PUBMED 7608087

COMMENT Related sequences: J01641 and M10196 (bases 1-1300); M77749 (bases 1366-2300); 3242-4542; Gene 33, 103, 1985 (bases 2300-2784); Gene 27, 183, 1984 (bases 2784-3242).
Location/Qualifiers

FEATURES
source 1..4542
/organism="synthetic construct"
/db_xref="taxon:32630"
/lab_host="Escherichia coli"
/plasmid="PBAD24"
96..974
/gene="arac"
96..974
/gene="arac"
1003..1020
/note="Operator O2"
1125..1153
/note="PC promoter"
1161..1182
/note="Operator O1"
1204..1217
/note="CAP site"
1213..1251
/note="Operator I2 + I1"
1250..1277
/note="PBAD promoter"
1300..1366
/note="multiple cloning site"
1306..1311
/note="Optimized RBS"
1317..1319
/note="Kozak sequence"
1320..1403
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAA57431.1"
/db_xref="GI:551425"
/translation="WVPGDPLESTCRHASLAVLADERRRFA"
1367..1792
/gene="rnb"
1367..1792
/gene="rnb"
1847..1853
/gene="rnb"
1847..1853
/note="bla P3 promoter"
1885..2748
/gene="bla"
1885..2748
/gene="bla"
2784..3242
/note="M13 origin"
3248..3945
/note="pBR322 origin"
BASE COUNT 1110 a 1191 c 1136 g 1105 t
ORIGIN

Query Match 11.3%; Score 186.2; DB 12; Length 4542;
Best Local Similarity 91.6%; Pred. No. 1e-29;
Matches 197; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 1439 TTGACTAGTCAAGTATGGAATTGCGAGTCGGTACCGGGGATCTCTAGAGTCGACCTG 1498
Db 1292 TTTTGGGCTAGCAGGAGATTCACCATGTCATCGGCGGATCCTCTAGAGTCGACCTG 1351

OY 1499 CAGCATGCAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCTGATACAGATT 1558
Db 1352 CAGCATGCAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCTGATACAGATT 1411

OY 1559 AAATCAGAACGCAAGAGCGGTCTGATAAACAAGATTTGCTGGCGGACATGAGCGCGTG 1618
Db 1412 AAATCAGAACGCAAGAGCGGTCTGATAAACAAGATTTGCTGGCGGACATGAGCGCGTG 1471

OY 1619 GTCCACCTGACCCCATGCGCACTCAAGTGA 1653
Db 1472 GTCCACCTGACCCCATGCGCACTCAAGTGA 1506

Search completed: January 16, 2003, 08:18:44
Job time : 2967 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 06:53:16 ; Search time 253 Seconds
(without alignments)
14713.649 Million cell updates/sec

Title: US-09-830-111A-1
Perfect score: 1653
Sequence: 1 ttttgggggtgaaaagtc.....atgcggaactcagaagtga 1653

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1653	100.0	1653	22	AAF74893
2	395.8	23.9	1353	24	ABN81605
3	242.8	14.7	1137	24	AAL40138
4	230.4	13.9	1212	24	AAL42942
5	230.4	13.9	1278	24	AAL42941
6	230.4	13.9	1614	24	AAL42940
7	220.4	13.3	1134	18	AA791893
8	198	12.0	3753	24	ABA04129
9	198	12.0	3755	24	ABA04130
					Saitoella complica
					Fungal decaprenyl
					Isoprenoid related
					Rhodotorula minuta
					Rhodotorula minuta
					Rhodotorula minuta
					Decaprenyl synthas
					Plasmid pHCE19(I)
					Plasmid pHCE19(II)

10	198	12.0	4176	10	AA90709	Sequence of plasmid
11	198	12.0	4189	11	AAQ05397	Secretion Vector p
12	198	12.0	4665	24	ABN84393	Antibiotic-indepen
13	195.6	11.8	1506	24	ABN81604	Fungal decaprenyl
14	187.2	11.3	1293	23	ABL03299	Drosophila melanog
15	180.8	10.9	966	21	AA661900	DNA encoding strai
16	180.8	10.9	966	24	ABQ62170	Arabidopsis sp. AT
17	180.8	10.9	966	24	AA596924	Arabidopsis locoph
18	179.2	10.8	5926	19	AAV32977	Tn7 donor plasmid
19	179.2	10.8	5926	19	AAV32978	Tn7 donor plasmid
20	170.4	10.3	2873	24	ABQ54840	Human ovarian anti
21	155.2	9.4	4009	15	AAQ70219	Plasmid pIG2 for T
22	152.4	9.2	3423	12	AAQ11083	Plasmid pIGRI for
23	152.4	9.2	3423	20	AAQ08615	pIGRI2 expression
24	152	9.2	1097	19	AAV39841	tac promoter/rnrb
25	152	9.2	1097	19	AAV31653	tac promoter/rnrb
26	152	9.2	1097	20	AAQ93172	P. fluorescens tac
27	148.8	9.0	6312	16	AAQ74814	Barley beta-amylase
28	148.8	9.0	6312	17	AAQ29193	Plasmid pBETA92/se
29	148.8	9.0	6312	17	AAQ03696	Plasmid pBETA92, e
30	147.8	8.9	5027	20	AAQ78865	Human tissue facto
31	147.8	8.9	5069	20	AAQ78874	Human tissue facto
32	147.8	8.9	5069	20	AAQ78871	Human tissue facto
33	147.8	8.9	5099	20	AAQ78896	Human tissue facto
34	147.8	8.9	5132	20	AAQ78883	Human tissue facto
35	147.8	8.9	5324	20	AAQ78890	Human tissue facto
36	147.8	8.9	5342	20	AAQ78893	Human tissue facto
37	147.6	8.9	6150	20	AAQ87427	Plasmid pBAD24-rec
38	147.2	8.9	518	19	AAV26002	Bacteriophage DNA
39	147.2	8.9	6806	24	ABL49925	Maltose binding prote
40	147	8.9	3474	13	AAQ29146	pMTNF-WPH plasmid
41	147	8.9	3474	14	AAQ51546	Expression plasmid
42	147	8.9	3950	22	AAQ81854	Plasmid p89-92 nuc
43	147	8.9	3950	24	ABL54456	Nucleotide sequenc
44	147	8.9	6553	21	AAQ55456	Destination vector
45	147	8.9	8430	21	AAQ50911	Recombinant human

ALIGNMENTS

RESULT 1

AAF74893 standard; DNA; 1653 BP.

AAF74893; AC

23-MAY-2001 (first entry)

Saitoella complicata decaprenyl diphosphate synthase encoding DNA SEQ:1.

Saitoella complicata; decaprenyl diphosphate synthase; coenzyme Q10;

fungus; ds.

Saitoella complicata.

Key Location/Qualifiers

CDS 134..1456

/*tag= a

/product= "decaprenyl diphosphate synthase"

MO200114567-Al.

01-MAR-2001.

24-AUG-2000; 2000WO-JP05659.

24-AUG-1999; 99JP-0237561.

(KANF) KANEKA CORP.

Matsuda H, Kawamukai M, Yajima K, Ikenaka Y, Hasegawa J;

Takahashi S;

XX WPI: 2001-202937/20.
DR P-PSDB: AAB74623.
XX DNA encoding a protein having decaprenyl diphosphate synthase activity
PT and microorganism for producing coenzyme Q10
XX
XX
XX Claim 1; Page 23-27; 32pp; Japanese.
XX
XX The present invention describes a method for microbiologically producing
CC coenzyme Q10 at a high efficiency by using a gene of the synthesis of
CC coenzyme Q10 side chain originating in a fungus belonging to the genus
CC *Saitoella*. The present sequence encodes the specifically claimed
CC *Saitoella* complicated protein having decaprenyl diphosphate synthase
CC activity. The protein having decaprenyl diphosphate synthase activity
CC can be used for producing coenzyme Q10.
XX
XX Sequence 1653 BP; 392 A; 417 C; 475 G; 369 T; 0 other;

Query Match 100.0%; Score 1653; DB 22; Length 1653;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGTGGGGTGAAGGCGGACGGGTCAGAGTTGGCTTGAGACGATAAGGCTCG 60
DB 1 TTTTGTGGGGTGAAGGCGGACGGGTCAGAGTTGGCTTGAGACGATAAGGCTCG 60
QY 61 GAGATTGAGTTGAGGACCAAGCTTTGATCGTGAAGTCTACATCTTCAGCAATATTTTC 120
DB 61 GAGATTGAGTTGAGGACCAAGCTTTGATCGTGAAGTCTACATCTTCAGCAATATTTTC 120
QY 121 AAATCATATATACATAGGCTCTACACAGACTGGGATACAGACATCAGCTTCGATCAT 180
DB 121 AAATCATATATACATAGGCTCTACACAGACTGGGATACAGACATCAGCTTCGATCAT 180
QY 181 CGCCCTCTGCGATCGGTTTACCCCTAAGAACAGCCTCGGACCTTCAATTGCACTAAGATG 240
DB 181 CGCCCTCTGCGATCGGTTTACCCCTAAGAACAGCCTCGGACCTTCAATTGCACTAAGATG 240
QY 241 TACCCCGAGACGCGCCCATCGAGTTGATGGCTGCTGTGTCTTCGCGCTCGAGACT 300
DB 241 TACCCCGAGACGCGCCCATCGAGTTGATGGCTGCTGTGTCTTCGCGCTCGAGACT 300
QY 301 GGTGAGCTGTGATCCGAATCAACTCTCTCATCATCCGCTCAACTTGGTCCGAGAT 360
DB 301 GGTGAGCTGTGATCCGAATCAACTCTCTCATCATCCGCTCAACTTGGTCCGAGAT 360
QY 361 GTCAATCTTACATCCACATCCGATCTCTCGGTTGAGACACCCCTTCTCGACAC 420
DB 361 GTCAATCTTACATCCACATCCGATCTCTCGGTTGAGACACCCCTTCTCGACAC 420
QY 421 TGTGCTAAATCTAATGTTGATGCTGAGGAAAGCATATTCGCTCGCTCATGTA 480
DB 421 TGTGCTAAATCTAATGTTGATGCTGAGGAAAGCATATTCGCTCGCTCATGTA 480
QY 481 GATGCTCAAGCGACGAGAGTTGCGCAAAAGTTCAAGGTTGGAGAGGTGTGAGAT 540
DB 481 GATGCTCAAGCGACGAGAGTTGCGCAAAAGTTCAAGGTTGGAGAGGTGTGAGAT 540
QY 541 TCCGGTGAACGAGGAGCTGACACCAAGAGTGTCAATGACAGAAACCCAGATATGAT 600
DB 541 TCCGGTGAACGAGGAGCTGACACCAAGAGTGTCAATGACAGAAACCCAGATATGAT 600
QY 601 GAACATGAGTGAAGCACTTAACGAAAGACGCGAGATCGAGGACAGACCTCGATAT 660
DB 601 GAACATGAGTGAAGCACTTAACGAAAGACGCGAGATCGAGGACAGACCTCGATAT 660
QY 661 CTTGCTTCGCAACGCGGTTGGTGAAGTCAAGAGATGATTCATGACATCACTCT 720
DB 661 CTTGCTTCGCAACGCGGTTGGTGAAGTCAAGAGATGATTCATGACATCACTCT 720
QY 721 CCAGACGACGTTATTCAGAGCTTCCGAGACGAGAAACGACCATCCGAAACACAGGC 780
DB 721 CCAGACGACGTTATTCAGAGCTTCCGAGACGAGAAACGACCATCCGAAACACAGGC 780

DB 721 CCAGACGACGTTATTCAGAGCTTCCGAGACGAGAAACGACCATCCGAAACACAGGC 780
QY 781 ATTGGAAACAGATGCGGATTTTGGCTGTGATTTCTTGTGGAGCGGCTGTGTC 840
DB 781 ATTGGAAACAGATGCGGATTTTGGCTGTGATTTCTTGTGGAGCGGCTGTGTC 840
QY 841 ATTGGCGAGTTGGCAATCCGAGAGTATGACTTTTGGCTGCTGATATGCAACTT 900
DB 841 ATTGGCGAGTTGGCAATCCGAGAGTATGACTTTTGGCTGCTGATATGCAACTT 900
QY 901 GGTGAGGAGATGTTACATGACAGTTGAAATAATCTGATGATGATGAGCTACGCG 960
DB 901 GGTGAGGAGATGTTACATGACAGTTGAAATAATCTGATGATGATGAGCTACGCG 960
QY 961 GACGAGAAACGTTGATCTAATTGACAGACATTAATTGAAAGCTGCTCTGAT 1020
DB 961 GACGAGAAACGTTGATCTAATTGACAGACATTAATTGAAAGCTGCTCTGAT 1020
QY 1021 TGGCAAGTCGACAGAGAACTGGGCTTGGGTGGTACAGCTGAGAGTTGCTGATGC 1080
DB 1021 TGGCAAGTCGACAGAGAACTGGGCTTGGGTGGTACAGCTGAGAGTTGCTGATGC 1080
QY 1081 TGTATGCTTACGGAAGAACTTGGTTGGCATTCAGATCGTCGACGATGCTCGA 1140
DB 1081 TGTATGCTTACGGAAGAACTTGGTTGGCATTCAGATCGTCGACGATGCTCGA 1140
QY 1141 CTACACGCTCTCCGCTACGACCTGGTAAACCGCGGTGCAAGCTCCAGCTCGGCT 1200
DB 1141 CTACACGCTCTCCGCTACGACCTGGTAAACCGCGGTGCAAGCTCCAGCTCGGCT 1200
QY 1201 CGCCACGCGCGGCGGCTTGGATGAGAACACACAGCGGAGCTCGGCTCCATGATCAA 1260
DB 1201 CGCCACGCGCGGCGGCTTGGATGAGAACACACAGCGGAGCTCGGCTCCATGATCAA 1260
QY 1261 GCGCAAGTTCTCTGACCCGAGAGAGCTGACGCGTGAACGAGTTGTGAGAAAGTGA 1320
DB 1261 GCGCAAGTTCTCTGACCCGAGAGAGCTGACGCGTGAACGAGTTGTGAGAAAGTGA 1320
QY 1321 TGGATTGAGAAACGAGAGCTTGGCGGAGAGATGCCCCAGAAAGCTTGGATGCAAT 1380
DB 1321 TGGATTGAGAAACGAGAGCTTGGCGGAGAGATGCCCCAGAAAGCTTGGATGCAAT 1380
QY 1381 TCGACGTTTCCCGAGAGTCCGCGACGGAAGCTTGGAGAGATGACGACAGAGCTT 1440
DB 1381 TCGACGTTTCCCGAGAGTCCGCGACGGAAGCTTGGAGAGATGACGACAGAGCTT 1440
QY 1441 GACTAGTCAAGATGAAATTCGAGCTCGTACCCGGGATCTCTAGAGTCAAGCTGCA 1500
DB 1441 GACTAGTCAAGATGAAATTCGAGCTCGTACCCGGGATCTCTAGAGTCAAGCTGCA 1500
QY 1501 GGCATGCAAGCTTGGCTGTTTGGCGGATGAGAAAGATTTTCAAGCTGATACAGATTA 1560
DB 1501 GGCATGCAAGCTTGGCTGTTTGGCGGATGAGAAAGATTTTCAAGCTGATACAGATTA 1560
QY 1561 ATCGAAGCGAGAGCGGTCTGATTAACAGATTTGCTGCGCGGAGTACCGGCTGCT 1620
DB 1561 ATCGAAGCGAGAGCGGTCTGATTAACAGATTTGCTGCGCGGAGTACCGGCTGCT 1620
QY 1621 CCCACCTGACCCCATGCGCAACTCAGAAAGTGA 1653
DB 1621 CCCACCTGACCCCATGCGCAACTCAGAAAGTGA 1653

RESULT 2
ABN81605
ID ABN81605 standard; DNA; 1353 BP.
XX
XX ABN81605;
XX
XX 29-AUG-2002 (first entry)
XX
XX Fungal decaprenyl diphosphate synthase encoding DNA SEQ ID NO 2.
XX


```
QY 387 CTCTCTCGGTTGAGGACACGCTTCTCTCGACACTGTCGCTAAATACTATGTTCACTCTG 446
Db 65 CCCTACTGGGATCAGGTATCATCCGCGCCCTAGACACGATAGCAAAAGTACTACTTCCAAAGCG 124
QY 447 AGGGAAGACATATTCGTCGCGCTCATGTGATGCTGATGGCTCAGGCGACGGA-GGTTGCG 505
Db 125 AGGCAACATATTCGCGCTATGATCGTTCTTCTCATGTCCCAAGCCACAAACGGTCTAG 184
QY 506 CCAAAAGTTTCAGGTTGGGAGAGGTCGTGGAGGTTCCGGTGAAACGAGGACTCGCACCA 565
Db 185 CGCCCGGTTTGAAGAACGCTCAAAATTTGAACTATCAGGTCGGAACACAGACTGATCCCT 244
QY 566 CCAG-----AGGTCTCAATGACAAAGAACCCAGATATGATGAACATGAGGTCAGGAC 617
Db 245 CCAGGTCAATCAATGATCTCTCGAAGTGAAGCAGATGAGATACTCAACGATTCGAATC 304
QY 618 CATTAAAGCAAGGACG-----GCGAGATCGAGGACAGACGTCGAATA 659
Db 305 CTTCTTGTTCGCTGCGAGCTCTCTTCGCGGCTCGATAGCATGCCGTCCACGTGCAATG 364
QY 660 TCCTCGCCTCGCAACGCGGTTGCTGAGATCACGGAGATGATCCATGACGATCACTCC 719
Db 365 TCCTACCTCGCAACGACGCTCGCGAAATCACCGAAATGATCCACGTAGCTTCGCTAT 424
QY 720 TCCACGACGAGTTATCGACGCTTCGAGACGACGACAAACGACCATCCGGAACCCAGG 779
Db 425 TGACGACGATGTCATAGACGTTTCAGCCATGAGGAGACACAAAGCGTCGCGCCCGCTG 484
QY 780 CATTTCGGAACCAAGATGGCGATTTTGGCTGTGATTTCTTTGGGACGGCGTCTGTTG 839
Db 485 CATTTCGGAACCAAGATCTCGGTCTGGCGGGGATTTCTCTCGCTCGTCTGCTGCTGT 544
QY 840 CATTTCGAGGTTTCGCAATCCGAGGTGATGAGCTTTTGGCTACTGTTATTGCAAACT 899
Db 545 ACCTCTCCGACTAGGAGCAACGAGGTCTGTCGAGCTAGTAGCTCGTCTGCTAGCTAATC 604
QY 900 TGGTTGAGGAGAGTTTATCGAGTTGAAATACTGTTGATGATGCGAAT-----949
Db 605 TAGTAGAGGCGAAGTCATGACATCAAGGAAATGCTCTGAAAGCAATGCAAGCGGAA 664
QY 950 -----GAGGCTACGCGACGAGGAAACGTTTCGATTACTATTTCGAGGA 995
Db 665 GCAAAGAGGTAGCAGTGCACAGATTGACCCCGGAAATTTTCGAACTATTATGAAGAAG 724
QY 996 CTACTTTGAAGACTGCTCTCTGATTCGCAAGTCGTGACAGCAAGTCGCTTCGGTG 1055
Db 725 CATACTTGAAGACCGCAAGTCTCATCGGAAATCGACAAGCGCACCTATCTCTCGGTG 784
QY 1056 GTGCTAGCCCTGA-----GGTTGCTGATGCTGCTTATGCTT 1091
Db 785 GAGCAGCGGAGAAACAGGGGTGATAGAGGCGAGCGCATATAAGACATTTGCGTACTCGT 844
QY 1092 ACCGAAGGACCTTGGTTTCGCTTCCAGATCGTCGACGACATGCTCGACTACACGCTCT 1151
Db 845 ACGTTCGCAATCTAGGTATTTGCTTTCCAGTCTGTCGACGATCTACTAGATTTCACAGCTA 904
QY 1152 CCGTACTCGGACCTCGGTAAGCCCG-----CGGTGACAGCTCCAGCTCGGTCTGCCACCG 1208
Db 905 CAGACGCGCAATTCGGCAAGCCCTCACAGGGTGCAGATCTGAAGCTCGGTCTCGCAACTG 964
QY 1209 CGCGCGCCCTTTTGGCATGGAAGCACACCGCGAGCTCGGTCCCATGATCAAGCGCAAGT 1268
Db 965 CGCGCGGCTGTACGATGGAGAGGTTCCCGGAGATGGCCAGATGATTTCTCGCAAGT 1024
QY 1269 TCTCTGACCCAGGAGAGCTCGAGGTGCACGCGAGTTGGTCGAGAAAGTATGATGGATGG 1328
Db 1025 TTGAGAACGAAGGCGGATGTTCGAAATCTCCAGGAATCTAGTAAGAAAGTCAAGTGGACCG 1084
QY 1329 AGAAGACGAGAGCCTTCGGCGGAGGATATCCCAAGAGGCGTTGGATGCAATTCGAGCGT 1388
Db 1085 AAAAGACCGTGAATTTGGCGGAAAAACATGCCGCACTCGCAATGGAGGCCCTCGAGGGAT 1144
```

```
QY 1389 TCCCGAGAGTCCGGCACGGAAGGCTTTGGAGCAGTTGACGCAAGGCTTCTACTAGGT 1448
Db 1145 TCCCGAGTCCGACGCTAGAGAAGCGCTCGAAGGCTGACCAAGACTGTGCTCAACCGAA 1204
QY 1449 CAAGATAG 1456
Db 1205 CAAAGTAG 1212

RESULT 5
AAL42941
ID AAL42941 standard; DNA; 1278 BP.
XX
AC AAL42941;
XX
DT 08-AUG-2002 (first entry)
XX
Rhodotorula minuta decaprenyl diphosphate synthase 2 coding sequence.
XX
Pungi; decaprenyl diphosphate synthase; gene; ds; enzyme;
XX high-yield fermentation; coenzyme Q10 production.
XX
OS Rhodotorula minuta.
XX
FH Key Location/Qualifiers
FT CDS 1..1278
/*tag= a
/*product= "Rhodotorula minuta decaprenyl diphosphate
synthase 2"
FT
XX
PN WO2002040682-A1.
XX
PD 23-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-JP10119.
XX
PR 20-NOV-2000; 2000JP-0352940.
XX
PA (KANF ) KANEKA CORP.
XX
PI Matsuda H, Kawamukai M, Yajima K, Ikenaka Y;
XX
DR WPI; 2002-427097/45.
DR P-PSDB; AAO14998.
XX
PT Decaprenyl diphosphate synthase gene of Rhodotorula origin for
PT efficient preparation of coenzyme Q10
PS Claim 2; Page 36-39; 50pp; Japanese.
XX
CC The invention comprises the amino acid and coding sequences of fungal
(Rhodotorula minuta) decaprenyl diphosphate synthase enzymes. The fungal
decaprenyl diphosphate synthase DNA and protein sequences are useful for
the efficient, high-yield fermentative production of coenzyme Q10 - for
pharmaceutical use. The present DNA sequence encodes a Rhodotorula minuta
decaprenyl diphosphate synthase enzyme.
XX
SQ Sequence 1278 BP; 345 A; 349 C; 334 G; 250 T; 0 other;

Query Match 13.9%; Score 230.4; DB 24; Length 1278;
Best Local Similarity 54.1%; Pred. No. 3e-49;
Matches 654; Conservative 0; Mismatches 476; Indels 78; Gaps 6;

QY 327 TCATCAATCCGCTCAACTTGGTCGGTCCGAGATGTCAAATCTTACATCCAAACATCCGAT 386
Db 71 TCTTCGACCACCTCCAACTCGTAGGCAATGAAGTCTCAAGTCTCCGGTCAACGTCGAAG 130
QY 387 CTCTCTCGGTTTCAGGACACCCCTTCTCTCGACACTGTCGCTAAATACTATGTTCACTCTG 446
Db 131 CCCTACTGGGATCAGGTATCTCCGCGCCCTAGACACGATAGCAAAAGTACTACTTCCAAAGCG 190
QY 447 AGGGAAGACATATTCGTCGCGCTCATGTGCTGCTGATGGCTCAGGCGACGGA-GGTTGCG 505
```

Db 191 AGGGAACATATTTGGGCTATGATCGTTCTTCATGTCACCAAGCACAACGGCTAG 250
 QY 506 CCAAAAGTTACAGGCTTGGGAGAAAGTCGTGAGGTTCCGGTGAACGAGGACTCGCACCA 565
 Db 251 CGCCCGGGGTTTGAAGAACGCTCAAAATTTGAACCTATCAGGTGCGAAAACAGACTGATCCCT 310
 QY 566 CCAG-----AGTGCTCAATGACACAGAACCCAGATATGATGAACTAGAGTCAAGAC 617
 Db 311 CCAAGTCAATCAATATCTCTCGAAGTGAAGACAGATGAGTACTCAACGATCGAATC 370
 QY 618 CATTAACGAGAGCG-----GCCAGATCCAGGAGCAAGAGCTCGAATA 659
 Db 371 CCTCTTCTGTCGCTCGAGACTCCTCTTGGCCGCTGATAGCATGCCGTCCAGCTCGAATG 430
 QY 660 TCCTGCGCTCGCAACGCGCGGTTGGCTGAGATCAAGGATGATGATGATGAGATCACTCC 719
 Db 431 TCCTACCTCGCAACGAGCGCTCGGGAATCAACGAAATGATCCAGTAGCTTCGCTAT 490
 QY 720 TCCACGACGACGTTATCCAGCGCTTCGAGACCAAGACGAAACGCAATCCGGAACCAAG 779
 Db 491 TGCACGACGATGTCATAGACGCTTCAGCCATGAGAGACAGACAGAGGTCGCCGCCGCTG 550
 QY 780 CATTCGGAACAAAGATGGGATTTGGCTGATGATTTCTTGTGGGACGGGCGTCTGTG 839
 Db 551 CATTCGGGAACAAAGATCTGCTGCTGGGCGGGAATTTCTCTCGCTCGCTTCGCTGT 610
 QY 840 CATTCGCGAGTTGGCGCAATCCGAGGTGATGATGATTTGGCTATCTGTTATTCGAAC 899
 Db 611 ACCTCTCCGAGCTAGGAGCAACGAGGTGCTGAGCTAGTACCTCGTGTAGCTAATC 670
 QY 900 TGGTGGAGGAGATTCATGACGTTGAAAAATATCTGTATGATGCGAAT----- 949
 Db 671 TAGTAGAGGCGCAAGTCATGACATCAAGGAAATGCTCTGAAAGCAATGCAAGCGGA 730
 QY 950 -----GAGGCTACGGCGACGAGGAAAGTTGATTAATTTGCGAAGA 995
 Db 731 GCAAAAGGTAGCAGTGCACAGATTGACCCCGAATTTTCGAACTTATATGAAAGAA 790
 QY 996 CTTACTTGAAGACTGCGTCTTGAATTCAGAGTCGTGACAGACAGTGGCTTCTGGGTG 1055
 Db 791 CATACTTGAAGACCGAATCTCATGCGAATTCGAAAGAGGACCACTATCTCGGTG 850
 QY 1056 GTGTAACGCTGA-----GGTGTGATGCTGCTTATGCTT 1091
 Db 851 GAGCAGGCGAGAAACAGGGGTGATAGAGGGGAGCGCATAAAGCATTTGGTACTCGT 910
 QY 1092 AGCGAAGAACTTGTGTTGGCATTCAGATCGTGAAGCAATGCTCGAATCAACCGTCT 1151
 Db 911 ACGGTGCAATCTAGGTATGCTTTCACACTGCTGACAGATCTACTAGATTTCACAGCTA 970
 QY 1152 CGGCTACGACCTCGGTAGACCCCGC---CGGTGACAGCTCCAGCTCGGCTCGCCACCG 1208
 Db 971 CAGAGCGCGAATTCGCGCAAGCCCTCAGAGGTGCAATCTGAAGCTCGGTCTCGCAACTG 1030
 QY 1209 CGCGGCGCTTTCGATGAAAGCACACACCGAGCTCGTCCATGATCAAGCGCAAT 1268
 Db 1031 CGCCGCGCTGTAGCATGGAAGATTCCCGAGATGGGCGAGATGATTTCTCCGCAAGT 1090
 QY 1269 TCTTCGACCCAGAGACCTGACGCTGACCGGAGTGGTCCAGAAAGTGAATGATGATGG 1328
 Db 1091 TTGAGAACCAAGCGCATGTCGAAATGCGAGAAATCTAATAGAAAGTCAAGCTGGACCG 1150
 QY 1329 AGAAGCGAGAGCTTGGCGGAGAGATATGCCAGAAAGCGTTGATGCAATTCGAGACT 1388
 Db 1151 AAAAAGCCTGAATTTGGCGGAAAAACATGCCCACTCCGATGAGGCCCTCGCAGGAT 1210
 QY 1389 TCCCGAGAGTCCGCGACCGAAAGCTTTGAGAGCATTTGACGAAAGGTTGACTAGGT 1448
 Db 1211 TCCCGAGTCCGAGCGCTAGAGAAAGCGCTGAAAGGCTGACCAAGACTGTGCTCAACGA 1270
 QY 1449 CAAAGTAG 1456
 Db 1271 CAAAGTAG 1278

RESULT 6
 AAL42940
 ID AAL42940 standard; DNA, 1614 BP.
 XX
 AC AAL42940;
 XX
 DT 08-AUG-2002 (first entry)
 XX
 DE Rhodotorula minuta decaprenyl diphosphate synthase 1 coding sequence.
 XX
 FM Fungi: decaprenyl diphosphate synthase; gene; ds; enzyme;
 KM high-yield fermentation; coenzyme Q10 production.
 XX
 OS Rhodotorula minuta.
 XX
 PH Key Location/Qualifiers
 FT CDS 1..1614
 FT /tag= a
 FT /product= "Rhodotorula minuta decaprenyl diphosphate
 FT synthase 1"
 XX
 PV WO200240682-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 20-NOV-2001; 2001WO-JP10119.
 XX
 PR 20-NOV-2000; 2000JP-0352940.
 XX
 PA (KANF) KANEKA CORP.
 XX
 PI Matsuda H, Kawamukai M, Yajima K, Ikenaka Y;
 XX
 DR WPI: 2002-427097/45.
 DR P-PSDB: AMO14997.
 XX
 PT Decaprenyl diphosphate synthase gene of Rhodotorula origin for
 PT efficient preparation of coenzyme Q10 -
 XX
 PS Claim 1; Page 29-33; 50pp; Japanese.
 XX
 CC The invention comprises the amino acid and coding sequences of fungal
 CC (Rhodotorula minuta) decaprenyl diphosphate synthase enzymes. The fungal
 CC decaprenyl diphosphate synthase DNA and protein sequences are useful for
 CC the efficient, high-yield fermentative production of coenzyme Q10 - for
 CC pharmaceutical use. The present DNA sequence encodes a Rhodotorula minuta
 CC decaprenyl diphosphate synthase enzyme.
 XX
 SQ Sequence 1614 BP; 432 A; 462 C; 377 G; 343 T; 0 other;
 XX
 Query Match 13.9%; Score 230.4; DB 24; Length 1614;
 Best Local Similarity 54.1%; Pred. No. 3.3e-48;
 Matches 654; Conservative 0; Mismatches 476; Indels 78; Gaps 6;
 QY 327 TCATCAATCCGCTCAACTTGGTCGCTCCGAGATGTCAAATCTTAATCAATCCGAT 386
 Db 407 TCTTGACCCCACTCCAATCTGAGGCAATGAACTGTAAGTCTCCGCTCAACGTCGAAG 466
 QY 387 CTCTCTCGGTTACGAGACACCTTCTCTGACACTGTCGTAATTAATCTATGTTCACTTG 446
 Db 467 CCTTACTGGGATTCAGGTATCCGCTTACGACAGATGCAAGATCACTTCCAAAGCGG 526
 QY 447 AGGGAACGATATTCGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 505
 Db 527 AGGGGAACATATTTGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 586
 QY 506 CCAAAAGTTACAGGCTTGGGAGAAAGTCGTGAGGTTCCGTTGAACGAGGACTCGCACCA 565
 Db 587 CGCCCGGGTTTGAAGAACGCTCAAAATTTGAACTATCAGGTGCGAAAACAGACTGATCCCT 646
 QY 566 CCAG-----AGTGCTCAATGACAAAGAACCCAGATATGATGAAATGAGGTGACGAC 617

```

Db 647 CAGGTCAATCAATGATCTCTCGAAGTGAAGCAGATGAGTACTCAACGATTCGAATC 706
QY 618 CATTACGAGAGGAGC-----CGAGATCGAGGACAGACGTCGGAATA 659
Db 707 CTTCTTGGTTCGCTGCGAGCTCTCTTCGCGCTCGATAGCATCCGCTCCACGTCGAATG 766
QY 660 TCCTCGCTCGCAACGCGGTTGGCTGAGATCAGGAGATGATCCATGCGACATCATCC 719
Db 767 TCCTACCTCGCAACGAGCGCTCGCGAAATCACCGAAATGATCCAGTACGTTTCGCTAT 826
QY 720 TCACACGACGAGTTATCGAGCTTCGAGAGCCAGACGAAAGCACCATCCGGAACAGG 779
Db 827 TGCACGAGATGTCATAGACGGTTCACCCATGAGGAGACACAGCGTCGCCGCCCTG 886
QY 780 CATTGGAACACAGATCGGATTTTGGCTGGTGAATTTCTTTGGGACGGGCGTCTGTTG 839
Db 887 CATTGCGGAACAAGATCTCGTGTGTCGGCGGATTTCTCTCGCTCGTCTCGCTGT 946
QY 840 CATTGGCGAGTTGCGCAATCCGAGAGTGATGAGCTTTTGGCTACTGTTATTGCAAACT 899
Db 947 ACCTCTCCGACTAGGAGGACACGAGCTCGTGGAGTAGTCCGCTCGCTAGCTAATC 1006
QY 900 TGGTTAGGAGAGTTCATGCACTGAAATAACTGTTGATGATGCGATT----- 949
Db 1007 TAGTAGAGGCGAAGTCATGCAGATCAAGGAATGCTCTGAAAGCAATGCAAGCGGAA 1066
QY 950 -----GAGGTACCGCGACGACGAGGAACGTTTCGATTACTATTTCGAGAAGA 995
Db 1067 GCAAAGAGTAGCAGTGCACAGATTGACCCCGGAAATTTTTCGAACTATATGAGAAGA 1126
QY 996 CTTACTTTGAAGATCGCTCTTGTATTCGAAAGTCGTCGAGCAAGTGCCTTCGGGTG 1055
Db 1127 CATACTTTGAAGCGCAAGTCTCATCGGAAATCGAAGAGCGCACCACTATCTCGGTG 1186
QY 1056 GTGCTACGCTGA-----GGTGTGCTGATGCTGCTTATGCTT 1091
Db 1187 GAGCAGCGGAACACAGGGTGGATAGAGGCGAGCGCATAAAGACATTCGCTACTCGT 1246
QY 1092 ACGGAAGAACCTTTGGTATTCAGATCGTCGACGACATGCTCGACTACACCGTCT 1151
Db 1247 ACGTTCGCAATCTAGGTATTGCTTTCCAGTCTGTCGACGATCTACTAGATTACAGCTA 1306
QY 1152 CCGTACCGACCTCGTAAAGCCG-----CGGTGAGACCTCCAGCTCGGTCTCGCCACG 1208
Db 1307 CAGACGCGCAATTCGCGAAGCCCTCACAGGCTGAGATCTGAAGCTCGGTCTCGCAACTG 1366
QY 1209 CGCGGCGCTCTTCGCATGGAAGCACACCGAGCTCGGTCCCATGATCAAGCGCAAGT 1268
Db 1367 CGCGCGCTGTACGATGGAGAGTTCGCGGAGATGGGCGCAGATGATTCGCGCAAGT 1426
QY 1269 TCTCTGACCCAGGAGAGCTCGAGCGTGCACGCGAGTTGGTTCGAGAAAAGTGATGGATTG 1328
Db 1427 TTGAGACGAAGCGATGTCGAACTCCAGGAATCTAGTAAGAAAGTCACTGAGCGCG 1486
QY 1329 AGAAGACGAGAGCTTCGCGGAGAGATGATCCCAAGAGCGTTGGATGCAATTGCGACGT 1388
Db 1487 AAAAGACCGTGAATTTGCGGAAAAAATGCGCACTCGCAATGAGGCGCTCGAGGAT 1546
QY 1389 TCCGAGAGTCCGCGACGGAAGCTTTGGAGCAGTTGACGACAGGTGTTGACTAGGT 1448
Db 1547 TCGCGGAGTCGAGCTAGAGAAGCGCTCGAAGCGCTGCAACAGACTGTGCTCAACCGAA 1606
QY 1449 CAAGATAG 1456
Db 1607 CAAGTAG 1614

```

RESULT 7
 AAT91893
 ID AAT91893 standard; DNA; 1134 BP.
 XX
 AC AAT91893;

```

XX 12-MAR-1998 (first entry)
XX Decaprenyl synthase gene.
XX Decaprenyl synthase gene; heterogenic ubiquinone production;
KW transferase; mitochondrial; yeast; hexaprenyl diphosphate;
KW modification; genetic engineering; ds.
XX Schizosaccharomyces pombe.
XX Key Location/Qualifiers
CDS 1..1134
FT /*tag= a
FT /product= Decaprenyl_synthase
XX JP09173076-A.
XX 08-JUL-1997.
XX 27-DEC-1995; 95JP-0351243.
XX 27-DEC-1995; 95JP-0351243.
XX (ALPH-) ALPHA SHOKUHIN KK.
XX WPI; 1997-397034/37.
XX P-PSDB; AAW30762.
XX New method for heterogenic ubiquinone production - allows the side
XX chain of ubiquinone to be freely modified by genetic engineering
XX methods
XX Claim 9; Page 6; 8pp; Japanese.
XX A new method has been developed for the formation of heterogenic prenyl
XX transferase. The method involves the fusing of a fragment containing a
XX transition sequence of mitochondrial Saccharomyces cerevisiae
XX hexaprenyl diphosphate synthase structural gene Coq1 to a heterogenic
XX prenyl transferase gene and expressing the resultant fused gene. The
XX present sequence encodes a specifically claimed decaprenyl synthase.
XX The method allows the side chain of ubiquinone to be freely modified
XX by genetic engineering methods.
XX Sequence 1134 BP; 328 A; 220 C; 244 G; 342 T; 0 other;
Query Match 13.3%; Score 220.4; DB 18; Length 1134;
Best Local Similarity 56.5%; Pred. No. 1.1e-46;
Matches 451; Conservative 0; Mismatches 341; Indels 6; Gaps 2;
QY 649 GAGCTCGAATATCTCGCTCGCAACGCGGTTGGCTGAGATCACGAGATGATCCATGC 708
Db 333 GACGGGTCAAATTTCTCCCTCTCAATTGAGATTAGCACAAATAACCGAGATGATCCATAT 392
QY 709 AGCATCACTCTCCACGACGCTTATCGACGCTTCGAGACCAGACGAAACGCCATC 768
Db 393 AGCAAGTTTGCTCGATGACGATGTGATTCATCAGCTTAATGTCGAGAGCTCACCTTC 452
QY 769 CGAAACACGAGCATTCGGAACAAAGATGCGGATTTTGGCTGGTGAATTTCTTTGGACG 828
Db 453 AAGCAATGTGTCTTCGTAATCGACCGCTCAATCTTCGCGGTAATTTTCATCTTGCAAG 512
QY 829 GGGCTGTGTTCGAGGTTGCGAGGTTGCGCAATCCGAGGTCATTGAGCTTTTGGCTACTGT 888
Db 513 GGCTTCGACTGTATGGCCGCTTCGAAATCCCAAGTTACGAGTTGTAGCTACAGT 572
QY 889 TATTGCAAACTTGGTTGAGGAGAGTTTCATGCGATTGAAAAAATCTGTTGATGATCGGAT 948
Db 573 GATAGCAGACTTGGTTTCGAGGTGAGTTTTCGAGCTAAAAAATCTATGATGATCCTTCATC 632
QY 949 TGAGGCTACGCGGACGAGGAAACGTTTCGATTACTATTTCAGAAAGACTTACTTTGAAGAC 1008
Db 633 TTTGGAAA---TAAAAACAATCAAATTTTGACTATTATTGAAAAAAGTTTGTGAA--- 686

```

QY 1009 TCGCTCTTGATTCGCAAGCTGTCGAGACGAGTGCCTTCTGGGTGGTCTACGCTCA 1068
DB 687 ACAGCGAGTTATTTCCAAAGGCTTCTACAACTCCGACCAATGTTCTCTAG 746
QY 1069 GGTTCCTGATGCTGCTTATGCTTACGAAAGAACCTGGTGGATTCAGATCGTCGA 1128
DB 747 TGTAGCAACAGCTGCTGGGAAATACGCTGATGATGATCTTCTTCAACTAATGGA 806
QY 1129 CGACATGCTGCACTACACCGTCTCCGCTACCGACCTCGGTAAACCCCGGATGACACT 1188
DB 807 TGACCTGTTGGACTATGATGCTGAAAGATGATCTTAGGAAAGCGGCTGGTGCAGATT 866
QY 1189 CCAGCTCGCTCGCCACCGCCGCTCTTCCGATGAAAGCACACCGCGAGCTCG 1248
DB 867 GAAGCTAGGGTGGCTACAGCTCCGCTCTTGTGATGAAAGATATCCAGAACTTGG 926
QY 1249 TCCCATGATCAAGCCGAGTCTGACCCAGAGACGTCGAGCGTGCACGCGAGTTGGT 1308
DB 927 TGCAATGATGTAATGAAATGATCAATCATCTTCTGATATCCAAAGGCTGGTCTTGGT 986
QY 1309 CGAGAAAGTGAATGATGAGAGAGACGAGCTTGGCGAGAGATATGCCAGAAAGC 1368
DB 987 TGAGTGCACATGATGCTATGAGCAAAACATCACTTGGGCAAAAGATATCAAAAAAGC 1046
QY 1369 GTTGATGCAATTCGAGCGTCCCGAGAGTCCGCGACGGAAGCTTTGGAGCATGAC 1428
DB 1047 CAAAGATTCCTTCTGTGCTCCCTGATTCACCTGCAAGAAAGACATTTTGGTGGC 1106
QY 1429 GACAAAGTGTGACTAG 1446
DB 1107 TGATTAAGTAATACGAG 1124

RESULT 8
ABA04129
ID ABA04129 standard; DNA; 3753 BP.
XX
AC ABA04129;
XX
DT 28-FEB-2002 (first entry)
XX
DE Plasmid pHCE19T(II) polynucleotide sequence SEQ ID NO:3.
XX
KW Promoter; gene expression; fermentation; D-AAT; Bacillus SK-1;
KW D-amino acid aminotransferase; economic; large-scale protein production;
KW circular; ds.
XX
OS Bacillus sp. SK-1.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT promoter 1..223
FT /tag= a
FT /note= "specifically claimed promoter region in
FT Claim 1 (see ABA04127)"
XX
PN WO200183787-A1.
XX
PD 08-NOV-2001.
XX
PF 26-APR-2001; 2001WO-JP03607.
XX
PR *27-APR-2000; 2000JP-0128528.
XX
PA (BIOL-) BIOLEADERS CORP.
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Sung M, Lee S, Hong S, Seo H;
XX
DR WPI; 2002-066535/09.
XX
PT Promoter sequence effective in Escherichia coli and Bacillus for

PT economic large-scale fermentative production of proteins
XX
XX Claim 7; Page 55-57; 65pp; Japanese.
XX
CC The present invention describes a DNA sequence (I) comprising promoter
CC activity in Escherichia coli or Bacillus cells, where the promoter is
CC derived from the D-AAT (D-amino acid aminotransferase) gene of
CC Bacillus SK-1. (I) can be used for the high level expression of a
CC foreign gene in a bacterial host for economic and efficient large-scale
CC production of proteins such as enzymes, cytokines and antibodies. The
CC present sequence represents the plasmid pHCE19T(II) polynucleotide
CC sequence comprising a specifically claimed promoter region from the
CC present invention.
XX
SQ Sequence 3753 BP; 953 A; 912 C; 929 G; 959 T; 0 other;
XX
Query Match 12.0%; Score 198; DB 24; Length 3753;
Best Local Similarity 100.0%; Pred. No. 9.4e-41;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1456 GGAATTCAGCTCGGTACCCGGGATCTCTAGAGTCGACCTGCAAGCATCAAGCTTGG 1515
DB 226 GGAATTCAGCTCGGTACCCGGGATCTCTAGAGTCGACCTGCAAGCATCAAGCTTGG 285
QY 1516 CTGTTTGGCGGATGAGAGATTTTCAGGCTGATACGATTAATCAGAACGACAGAG 1575
DB 286 CTGTTTGGCGGATGAGAGATTTTCAGGCTGATACGATTAATCAGAACGACAGAG 345
QY 1576 CGGTCTGATTAACGAAATTTGCTGGCGGAGTAGCGGGTGGTCCCACTGACCCCAT 1635
DB 346 CGGTCTGATTAACGAAATTTGCTGGCGGAGTAGCGGGTGGTCCCACTGACCCCAT 405
QY 1636 GCCGAATCTCAGAGTGAA 1653
DB 406 GCCGAATCTCAGAGTGAA 423

RESULT 9
ABA04130
ID ABA04130 standard; DNA; 3755 BP.
XX
AC ABA04130;
XX
DT 28-FEB-2002 (first entry)
XX
DE Plasmid pHCE19(II) polynucleotide sequence SEQ ID NO:4.
XX
KW Promoter; gene expression; fermentation; D-AAT; Bacillus SK-1;
KW D-amino acid aminotransferase; economic; large-scale protein production;
KW circular; ds.
XX
OS Bacillus sp. SK-1.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT promoter 1..225
FT /tag= a
FT /note= "specifically claimed promoter region in
FT Claim 1 (see ABA04128)"
XX
PN WO200183787-A1.
XX
PD 08-NOV-2001.
XX
PF 26-APR-2001; 2001WO-JP03607.
XX
PR *27-APR-2000; 2000JP-0128528.
XX
PA (BIOL-) BIOLEADERS CORP.
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Sung M, Lee S, Hong S, Seo H;
XX
PT Promoter sequence effective in Escherichia coli and Bacillus for


```
XX 21-JAN-1989; 89DE-3901681.
PR (BEHW ) BEHRINGWERKE AG.
XX
XX Knapp S, Abel KJ;
XX WPI; 1990-232260/31.
DR P-PSDB; AAR96222.
XX
XX Signal peptide from Bordetella pertussis - causing secretion of
PT heterologous proteins in E.coli, and expression vectors for isolating
PT and testing signal sequences.
XX
XX Disclosure; ; 8pp; German.
XX
XX Complete sequence of pSEC-BP1. Constructed from a 3.1kb SacI-ScaI
CC fragment of pTRC99C-phoB-seq1 (including signal sequence) ligated to
CC a 0.9kb SacI-ScaI fragment from pTRC97A. DNA encoding an
CC heterologous protein can be inserted into the polylinker region.
CC Upstream of this is a strong promoter (perfect match to consensus
CC -35 and -10 regions) and the B.pertussis signal sequence.
CC Downstream are three stop codons, one in each possible reading
CC frame.
CC See also AA005398-Q05400 and AA005521-Q05526.
XX
XX Sequence 4189 BP; 1004 A; 1067 C; 1123 G; 995 T; 0 other;
SQ
Query Match 12.0%; Score 198; DB 11; Length 4189;
Best Local Similarity 100.0%; Pred. No. 9.8e-41;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1456 GGAATTTCGAGCTCGGTACCCGGGGATCCTTAGAGTCGACCTGCAGCGCATGCAAGCTTGG 1515
DB 332 GGAATTTCGAGCTCGGTACCCGGGGATCCTTAGAGTCGACCTGCAGCGCATGCAAGCTTGG 391
QY 1516 CTGTTTTGGCGGATGAGAGAAATTTTCAGCTGATACGATTAATTCGAACGCAGAAAG 1575
DB 392 CTGTTTTGGCGGATGAGAGAAATTTTCAGCTGATACGATTAATTCGAACGCAGAAAG 451
QY 1576 CGGTCGTATAAACAAGAAATTTGCTGCGCGGAGTAGCGGCTGTCGCCACCTGACCCCAT 1635
DB 452 CGGTCGTATAAACAAGAAATTTGCTGCGCGGAGTAGCGGCTGTCGCCACCTGACCCCAT 511
QY 1636 GCCGAACCTCAGAAAGTGAA 1653
DB 512 GCCGAACCTCAGAAAGTGAA 529
XX
RESULT 12
ABN84393
ID ABN84393 standard; DNA; 4665 BP.
XX
XX ABN84393;
XX
XX 01-OCT-2002 (first entry)
XX
XX Antibiotic-independent high expression vector pHCE(III).
XX
XX Vector, pHCE(III); glutamate racemase; enzyme; selectable marker;
KW gene; ds.
XX
XX Chimeric - Bacillus borestelensis.
OS Chimeric - Unidentified.
XX
XX WO200255716-A1.
XX
XX 18-JUL-2002.
XX
XX 11-JAN-2002; 2002MO-KR00048.
XX
XX 12-JAN-2001; 2001KR-0001690.
XX
```

```
PA (BIOL-) BIOLEADERS CORP.
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
XX Sung MH, Lee SG, Hong SP, Yoon EJ, Choi YH, Poo HR;
XX WPI; 2002-557828/59.
XX
XX New antibiotic-independent plasmid vector, useful for stably expressing
PT and transcribing genes without the need for any antibiotic-resistant
XX genes, and for producing other recombinant proteins or nucleic acids -
XX
XX Claim 8; Page 41-44; 45pp; English.
XX
XX The present sequence is that of novel antibiotic-independent high
CC expression vector plasmid pHCE(III), deposited in Escherichia coli
CC as KCYC 0925 BP. The vector was constructed by insertion of the
CC Bacillus borestelensis glutamate racemase (GluA) gene (see ABN84392)
CC into vector pHCE19T(III). The GluA gene, being essential to the
CC microbial survival, was utilised as selectable marker. A recombinant
CC gene can be stably transcribed and highly expressed using this
CC vector, without the use of antibiotic resistance genes. Production
CC of GluA from the vector is sufficient to provide D-glutamic acid
CC for cell growth of a transformed host. The stability of the vector
CC was confirmed using pHCE(III)-TPI, which includes a tyrosine
CC phenol lyase gene, by transfer to E. coli WM335 (a D-glutamate
CC auxotroph) host cells and cultivation through 9 generations. When
CC expression vector pHCE(III)-TNA, which includes the Symbiodacterium
CC toebii SC-1 tryptophan indole lyase (TNA) gene, was transformed
CC into E. coli WM335 and cultivated, TNA production increased with
CC cell growth and a high expression level was maintained throughout
CC the growth period. The TNA content reached about 40% of total cell
CC protein. The vector resolves problems associated with the use of
CC expensive inducers and antibiotic-resistance genes. It can be
CC widely applied for large-scale production of exogenous proteins,
CC with no adverse effect on the environment.
XX
XX Sequence 4665 BP; 1141 A; 1193 C; 1145 G; 1186 T; 0 other;
SQ
Query Match 12.0%; Score 198; DB 24; Length 4665;
Best Local Similarity 100.0%; Pred. No. 1e-40;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1456 GGAATTTCAGCTCGGTACCCGGGGATCCTTAGAGTCGACCTGCAGGATCAGAGCTTGG 1515
DB 226 GGAATTTCAGCTCGGTACCCGGGGATCCTTAGAGTCGACCTGCAGGATCAGAGCTTGG 285
QY 1516 CTGTTTTGGCGGATGAGAGAAATTTTCAGCTGATACGATTAATTCGAACGCAGAAAG 1575
DB 286 CTGTTTTGGCGGATGAGAGAAATTTTCAGCTGATACGATTAATTCGAACGCAGAAAG 345
QY 1576 CGGTCGTATAAACAAGAAATTTGCTGCGCGGAGTAGCGGCTGTCGCCACCTGACCCCAT 1635
DB 346 CGGTCGTATAAACAAGAAATTTGCTGCGCGGAGTAGCGGCTGTCGCCACCTGACCCCAT 405
QY 1636 GCCGAACCTCAGAAAGTGAA 1653
DB 406 GCCGAACCTCAGAAAGTGAA 423
XX
RESULT 13
ABN81604
ID ABN81604 standard; DNA; 1506 BP.
XX
XX ABN81604;
XX
XX 29-AUG-2002 (first entry)
XX
XX Fungal decaprenyl diphosphate synthase encoding DNA SEQ ID NO 1.
DE
XX Fungi; Aspergillus; Leucosporidium; coenzyme Q10; enzyme;
KW decaprenyl diphosphate synthase; gene; ds.
XX
XX Leucosporidium scottii.
OS
```


CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB13511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 CC
 XX

Sequence 1293 BP; 321 A; 375 C; 346 G; 251 T; 0 other;

Query March 11.3%; Score 187.2; DB 23; Length 1293;
 Best Local Similarity 54.1%; Pred. No. 4e-38; Indels 21; Gaps 1;
 Matches 418; Conservative 0; Mismatches 333;

671 CAACGGCGGTTGGCTGAGATCAAGAGATGATCATGACATCACTCCTCCAGCAGAC 730
 Db 529 CAGCGACAGATGCGCTCTTTTCGGAGATGTCATCGGCCAGTGTGTCACGCGAT 588
 Qy 731 GTTATGACGCTTCCGAGACAGACGAAAGCAACATCCGGAACCGGATTCGGAAC 790
 Db 589 GTCATCGATCACTCGGACTTCCGACGGCGCAAGCCAGCGTAATGCTGTGCAACAC 648
 Qy 791 AAGATGGCGATTTGGCTGCTGATTTCTTTGGGACGGGCTGTGTCATTTGGCGAGG 850
 Db 649 AAGAAGGTCAACATGCTGTGATTAATCATCTTATGCTGCTGATTAATGATCTGCT 708
 Qy 851 TTGCGCAATCCGAGAGTATGAGCTTTTGGCTACTGTTATGCAAACTTGGTTAGGGA 910
 Db 709 CTGCCAGCGATGATGTCATGATCGTGTGATGATCATTTTATCCGATTTGGTCCAA 768
 Qy 911 GAGTTTCATGCACTGAAAAATACTGTGTATGATGCGATTGAGGCTACGGCGACGAG 970
 Db 769 GAGTTTCATGCACTGAGGCTCAAGGAAACGCA-----GAAACGAG 807
 Qy 971 AGCTTCGATTAATTTTGGAGAAAGTACTTGAAGACTGCGTCTTGTATTTGCCAAGTG 1030
 Db 808 CGCTTCGCCCTTATCTGCAACAGATACAGAAAGCCGATCGTGTATCCCAATGCA 867
 Qy 1031 TGACAGCAAGTGGCTTGTGGGTGTGTACGCCCTGAGAGTGTGATCTCTTATGCT 1090
 Db 868 CTGAAGGCGACCGCGTGTATTTCCAGCGACGACAGACGTGGCCGAGGTGGCTTCCAG 927
 Qy 1091 TACGGAAGAACTTGTGTGTCATTTCCAGATGTCGACGACATGCTGCACTACACCG 1150
 Db 928 TACGGAAGAACTGCGGTGCTTTCAGTGTGTCGACGACATGCTGCACTTGTGCTCC 987
 Qy 1151 TCCGTTACCGACTCGTGAAGCCCGCGGTGCAAGCTTCAAGTGTGCTGCGCACCG 1210
 Db 988 TCCACCGAGCAATGCGGCAAGCCGCGGCGGATTTGAAGCTGGTCTGCGCACCGCT 1047
 Qy 1211 CCGGCGCTTCTTGGCATGGAAGACCAACCGCGAGCTCGGCTCCCATGATCAAGGCA 1270
 Db 1048 CCGGCTCTTCTTGGCATGGAAGATACCCCGAGCTGAATCCCATGTGATGCGCGGCTTC 1107
 Qy 1271 TCTGACCCAGAGAGCTGAGCGGTGCAAGCTTGTGAGAAAAAGTATGATTTGAG 1330
 Db 1108 AGCGAGCCCGGAGAGCTGAGCGGCTTTCAGTGTGTCACAAAGCGACGCTGTGAG 1167
 Qy 1331 AAGACGAGAGCTTGGCGGAGAGATGTCGCAAGAGGCTTGGATGCAATTTGGAC 1390
 Db 1168 CAGACCCGGTCTTCTGCGCCAAAGACATGCAACGAGGCGATTCGCGTCCAGAGAG 1227
 Qy 1391 CCGGAGAGTCCGCGCAACGAAAGCTTGGAGCATTTGAGCGGACAGAGTGTGA 1442
 Db 1228 ACGAGTGCCTTACCAAGAGGCTTCCAGGTGTGCGGACTTGTATCA 1279

RESULT 15
 AAC61900
 ID AAC61900 standard; DNA; 966 BP.
 XX
 AC AAC61900;

XX 06-MAR-2001 (first entry)
 DT
 XX DNA encoding straight-chain class prenyltransferase designated ATPT8.
 XX
 DE
 XX Prenyltransferase; ATPT1; ATPT2; ATPT3; ATPT4; ATPT5; ATPT6; ATPT7;
 KW ATPT8; ATPT9; ATPT10; ATPT11; ATPT12; tocopherol; homogentisic acid;
 KW phytylpyrophosphate; 2-methyl-6phytylbenzoquinol; antioxidant;
 KW nutritional supplement; ss.
 KW
 XX Arabidopsis sp.
 OS
 XX

XX Key Location/Qualifiers
 FT 1..966
 FT CDS /*tag= a
 FT /product= "prenyltransferase ATPT8"

XX MO200063391-A2.

XX 26-OCT-2000.

XX 14-APR-2000; 2000MO-US10368.

XX 15-APR-1999; 99US-0129899.

XX 30-JUL-1999; 99US-0146461.

XX (CALJ) CALGENE LLC.

XX Savidge B, Lassner MW, Weiss JD, Post-Beltemmiller D;
 PI
 XX WPI; 2000-647519/62.

XX P-PSDB; AAB19412.

XX An isolated nucleic acid sequence encoding prenyltransferase used to
 PT transform plant cells to increase the production of tocopherols -
 XX

XX Example 1; Page 82; 114pp; English.

XX The present sequence encodes a prenyltransferase. The specification
 CC describes prenyltransferases designated ATPT1, ATPT2, ATPT3, ATPT4,
 CC ATPT5, ATPT6, ATPT7, ATPT8, ATPT9, ATPT10, ATPT11, and ATPT12. The
 CC biosynthesis of alpha-tocopherol in higher plants involves the
 CC condensation of homogentisic acid and phytylpyrophosphate to form
 CC 2-methyl-6phytylbenzoquinol, which can form various tocopherols. The
 CC prenyltransferase polynucleotides are useful in transforming host cells
 CC to alter the expression of prenyltransferase in these cells. The
 CC transformed cells are used in the production of tocopherols which are
 CC of use in the pharmaceutical industry as antioxidants and also in the
 CC food industry as nutritional supplements.

XX Sequence 966 BP; 278 A; 196 C; 240 G; 252 T; 0 other;

Query March 10.9%; Score 180.8; DB 21; Length 966;
 Best Local Similarity 54.1%; Pred. No. 1.6e-36;
 Matches 405; Conservative 0; Mismatches 322; Indels 21; Gaps 1;

Qy 650 ACGTGAATATCTCGCTCCGCAACGCGGCTGCTGATGATCAGAGATGATCAGCA 709
 Db 163 ACATCAGAAATTAACGGTAAAGCAACGGGATTTGCTGAAATCACTGAATATATACGTC 222
 Qy 710 GCATCACTCTCCACGACGACGATATGACGCTTCCGAGACCAAGCAACGACCATCC 769
 Db 223 GCAAGTCTACTGACAGATGATGCTTGGATGATCCGATACAAAGCGCGTGTGTTCC 282
 Qy 770 GGAACACGAGATTCGAAAGAAAGATGCGATTTTGGCTGTGATTTCTTGTGGAGCG 829
 Db 283 TTAATGTTGTAATGGGTAAACAGATGCGGTATTAAGCAGAGACTTCTTCCCGG 342
 Qy GCGTCTGTTGATTTGGCGAGAGTTGCGCAATCCGAGAGTATGAGCTTTTGGTACTGTT 889
 Db 343 GCTTGTGGGCTCTCGCTCTTTAAAGAACACAGAGTTGATGATTAATCTTGAACACTGCT 402
 Qy 890 ATTGCAAACTTGTTGAGGAGAGAGTTCATGCACTTGAATAAACTGTTGATGATGCGATT 949

Db 403 GTAGACATCTTTGTTACCGGTGAACCATGGAATTAAGTACTAGT----- 444
QY 950 GAGGCTACGGCGACGCGAGAAACGTTGATTTACTATTTCAGAAAGACTTACTTTGAAGACT 1009
Db 445 ---TCAACCGAGCAGCGTTATAGTATGACTACTACATGCAGAAGACATATTATAAGACA 501
QY 1010 GCGTCTTGTGATTCGCAAGTCGTCAGAGCAAGTCGCTTCTGGGTGGTGTACGCTGAG 1069
Db 502 GCATCGCTAATCTTAACAGCTGCAAGCTGTTGCCGTTCTCACTGGACAACACAGCAGAA 561
QY 1070 GTTGCTGATCTGCTATGCTTACGGAAGAACCTTGGTTTGGCATTTCCAGATCGTCGAC 1129
Db 562 GTTGCGGTGTAGCTTTTGAATATGGAGGAATCTGGTTTACATTTCCAAATTAATAGAC 621
QY 1130 GACATGCTGACTACACCGTCTCCGTACCGACCTCGGTAAGCCCGCGGTGCAGACCTC 1189
Db 622 GACATTTCTTGATTTACGGGCACATCTGCCTCTCTCGGAAGGGATCGTTGTTCAGATATT 681
QY 1190 CAGCTCGGTCTGCCACCGCGCGCCCTCTTCGCATGGGAAGCACACCGCGAGCTCGGT 1249
Db 682 CGCCATGGAGTCAATAACAGCCCCAATCTCTTTGCCATGGAGAGTTTCTCAACTACGC 741
QY 1250 CCCATGATCAAGCGCAAGTTCTCTGACCCAGAGAGCTGAGCGTGCACGCGAGTTGGTC 1309
Db 742 GAAGTTGTTGATCAAGTTGAAGAAGATCCTAGGAATGTTGACATTTGCTTTAGAGTATCTT 801
QY 1310 GAGAAAAGTGATGGATTGGAGAAAGACGAGAGCCCTTGGCGGAGGAGTATGCCCAAGAGCGG 1369
Db 802 GGGAGAGCAAGGGGAATACAGAGGGCAAGAGAATTAGCCATGGAACATGCGAATCTAGCA 861
QY 1370 TTGGATGCAATTCGGAGCTTCCCGGAGA 1397
Db 862 GCAGCTGCAATCGGCTCTCTACCTGAAA 889

Search completed: January 16, 2003, 08:23:14
Job time : 277 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 07:27:51 ; Search time 58 Seconds
(without alignments)
8740.288 Million cell updates/sec

Title: US-09-830-111A-1
Perfect score: 1653
Sequence: 1 ttttggggtcgaaagtc.....atcgcgaaactcagaagtga 1653

Scoring table: IDENTITY NUC

Gapop 10.0 , Capext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/prodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/prodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/prodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/prodata/1/ina/PCUTUS_COMB.seq.*
- 6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	179.2	10.8	5926	4	US-09-027-169-3
C 2	179.2	10.8	5926	4	US-09-027-169-4
C 3	160.4	9.7	3301	2	US-08-447-430A-42
4	155.2	9.4	4009	2	US-08-500-860A-2
5	152.4	9.2	3423	2	US-08-447-430A-40
6	152.4	9.2	3423	4	US-09-485-737B-86
7	152	9.2	1097	1	US-08-761-258-7
8	152	9.2	1097	2	US-08-977-306-7
9	148.8	9.0	6312	1	US-08-531-601-3
10	148.8	9.0	6312	2	US-08-859-032-3
11	147.2	8.9	518	4	US-09-242-901-3
12	147	8.9	3474	2	US-08-447-430A-41
13	147	8.9	3474	2	US-08-318-837-10
14	147	8.9	3474	4	US-08-122-458B-8
C 15	147	8.9	8430	4	US-09-131-028A-6
16	147	8.9	8430	4	US-09-131-028A-10
17	146.2	8.8	628	4	US-09-242-901-1
18	146	8.8	4476	3	US-08-801-344-2
19	146	8.8	4476	4	US-09-498-599-2
20	145.6	8.8	7475	2	US-08-971-036-1
21	145.6	8.8	7475	3	US-09-096-570-1
22	145.6	8.8	7475	4	US-09-265-617B-1
C 23	145.4	8.8	1200	4	US-09-204-117B-2
C 24	145.4	8.8	2650	4	US-09-204-117B-3
C 25	145.4	8.8	4199	4	US-09-204-117B-1
26	144	8.7	5096	1	US-08-106-433A-3
27	144	8.7	5110	1	US-08-106-433A-4

28	144	8.7	5110	1	US-08-106-433A-5	Sequence 5, Appli
29	144	8.7	5110	1	US-08-106-433A-6	Sequence 6, Appli
30	144	8.7	5110	1	US-08-106-433A-7	Sequence 7, Appli
31	144	8.7	5110	1	US-08-106-433A-8	Sequence 8, Appli
32	144	8.7	5110	1	US-08-106-433A-9	Sequence 9, Appli
33	144	8.7	5110	1	US-08-106-433A-10	Sequence 10, Appli
34	144	8.7	5110	1	US-08-106-433A-11	Sequence 11, Appli
35	144	8.7	5110	1	US-08-106-433A-12	Sequence 12, Appli
36	144	8.7	5110	1	US-08-106-433A-13	Sequence 13, Appli
37	144	8.7	5110	1	US-08-106-433A-14	Sequence 14, Appli
38	144	8.7	5110	1	US-08-106-433A-15	Sequence 15, Appli
39	138	8.3	5238	6	5453363-1	Patent No. 5453363
C 40	135.4	8.2	4593	3	US-08-801-344-1	Sequence 1, Appli
C 41	135.4	8.2	4593	4	US-09-498-599-1	Sequence 1, Appli
42	129.8	7.9	469	2	US-08-647-368A-1	Sequence 1, Appli
43	120.4	7.3	1464	4	US-09-276-531-18	Sequence 18, Appli
44	92	5.6	2263	2	US-08-687-865A-1	Sequence 1, Appli
45	92	5.6	2263	4	US-09-043-711-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-027-169-3/c
; Sequence 3, Application US/09027169
; Patent No. 6420524
; GENERAL INFORMATION:
; APPLICANT: CRAIG, NANCY L
; TITLE OF INVENTION: GAIN OF FUNCTION MUTATIONS IN
; NUMBER OF INVENTION: ATP-DEPENDENT TRANSDUCTION PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Anne Brown (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave.
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; ZIP: 27608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,169
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 5789-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2205
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5926 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "PEM delta R.adj to 1"
US-09-027-169-3

Query Match 10.8%; Score 179.2; DB 4; Length 5926;
Best Local Similarity 91.3%; Pred. No. 1.2e-37;
Matches 190; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1446 GGTCAAGATAGGAATTCGAGTCGGTACCGGGGATCCTCTAGAGTCGACCTCGAGCAT 1505
DB 4282 GTTATTAAAGCTGTCGACGAGCTTCTTTGGGGATCCTCTAGAGTCGACCTCGAGCAT 4223

QY 1506 GCAAGCTTGGCTGTTTGGCGGATGAGAGATTTTACGCTGATACAGATTAAATCAG 1565
DB 4222 GCAAGCTTGGCTGTTTGGCGGATGAGAGATTTTACGCTGATACAGATTAAATCAG 4163
QY 1566 AACGCAAGACGGCTGTGATTAACAGAAATTTGCTGGCGGACAGTACGGCGGTGCTCCAC 1625
DB 4162 AACGCAAGACGGCTGTGATTAACAGAAATTTGCTGGCGGACAGTACGGCGGTGCTCCAC 4103
QY 1626 CTGACCCCATGCCGAAGTCAAGAGTGA 1653
DB 4102 CTGACCCCATGCCGAAGTCAAGAGTGA 4075

RESULT 2
US-09-169-4/c
Sequence 4, Application US/09027169
Patent No. 6420524
GENERAL INFORMATION:
APPLICANT: CRAIG, NANCY L
TITLE OF INVENTION: GAIN OF FUNCTION MUTATIONS IN
TITLE OF INVENTION: ATP-DEPENDENT TRANSDUCTION PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Brown (Alston & Bird, LLP)
STREET: 3605 Glenwood Ave.
CITY: Raleigh
STATE: NC
COUNTRY: USA
ZIP: 27608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/027,169
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Anne
REGISTRATION NUMBER: 36,463
REFERENCE/DOCKET NUMBER: 5789-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2205
TELEFAX: 919 881 3175
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5926 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "pEM-delta"
US-09-027-169-4

Query Match 10.8%; Score 179.2; DB 4; Length 5926;
Best Local Similarity 91.3%; Pred. No. 1.2e-37;
Matches 190; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1446 GGTCAAGATGAGATTTGAGCTGCGTACCCGGGATCTCTAGAGTGCAGCTGAGGCAT 1505
DB 3918 GTTATTAAAGCTGTCCGACGGACTTCTTGGGGATCTCTAGAGTGCAGCTGAGGCAT 3859
QY 1506 GCAAGCTTGGCTGTTTGGCGGATGAGAGATTTTACGCTGATACAGATTAAATCAG 1565
DB 3858 GCAAGCTTGGCTGTTTGGCGGATGAGAGATTTTACGCTGATACAGATTAAATCAG 3799
QY 1566 AACGCAAGACGGCTGTGATTAACAGAAATTTGCTGGCGGACAGTACGGCGGTGCTCCAC 1625
DB 3798 AACGCAAGACGGCTGTGATTAACAGAAATTTGCTGGCGGACAGTACGGCGGTGCTCCAC 3739
QY 1626 CTGACCCCATGCCGAAGTCAAGAGTGA 1653

DB 3738 CTGACCCCATGCCGAAGTCAAGAGTGA 3711

RESULT 3
US-08-447-430A-42
Sequence 42, Application US/08447430A
Patent No. 5916558
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Recombinant polypeptides and peptides,
TITLE OF INVENTION: nucleic acids coding for the same and use of these
TITLE OF INVENTION: polypeptides and peptides in the diagnosis of
TITLE OF INVENTION: tuberculosis.
NUMBER OF SEQUENCES: 43
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,430A
FILING DATE:
CLASSIFICATION: 424
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 3301 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: plasmid vector
HYPOTHETICAL: NO
US-08-447-430A-42

Query Match 9.7%; Score 160.4; DB 2; Length 3301;
Best Local Similarity 97.8%; Pred. No. 9e-33;
Matches 174; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1476 GGGAGCTCTGAGCTGACCTGCGAGGATGCAAGCTGTTTGGCGGATGAGAGA 1535
DB 237 GGGAGCTCTGAGCTGACCTGCGAGGATGCAAGCTGTTTGGCGGATGAGAGA 293
QY 1536 AGATTTTCCGCTGATGATGATTAATCAGAACGAGAGCGGTGTGATTAACAGATT 1595
DB 294 AGATTTTCCGCTGATGATGATTAATCAGAACGAGAGCGGTGTGATTAACAGATT 353
QY 1596 TGCCTGGCGGAGTACGGCGGTGTCCACCTGACCCCATGCCGAAGTGA 1653
DB 354 TGCCTGGCGGAGTACGGCGGTGTCCACCTGACCCCATGCCGAAGTGA 411

RESULT 4
US-08-500-860A-2
Sequence 2, Application US/08500860A
Patent No. 5891679
GENERAL INFORMATION:
APPLICANT: LUCAS, RUDOLPH
APPLICANT: DE BAETSELIER, PATRICK
APPLICANT: FRANSSEN, LUCIE
APPLICANT: SABON, ERWIN
TITLE OF INVENTION: TYP-MUTATIONS, A PROCESS FOR PREPARING THEM AND
TITLE OF INVENTION: THEIR USE AS ACTIVE SUBSTANCES IN PHARMACEUTICAL COMPOSITIONS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/500,860A

FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-8

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)816-4000

TELEFAX: (703)816-4100

TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 4009 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

IMMEDIATE SOURCE:

CLONE: pig2

US-08-500-860A-2

Query Match 9.4%; Score 155.2; DB 2; Length 4009;
Best Local Similarity 96.6%; Pred. No. 2.3e-31;
Matches 170; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 1478 GGATCCTCTAGATCGACCTGCGGATGCAAGCTTGGCTGTTTGGCGGATGAGAG 1537

Db 947 GAATTTCTAGATCGACCTGCGGATGCAAGCTTGGCTGTTTGGCGGATGAGAG 1003

QY 1538 ATTTTCAGCTGTATACAGATTAATCAGAACGCGGCTCTGTATAAAACAGAAATTG 1597

Db 1004 ATTTTCAGCTGTATACAGATTAATCAGAACGCGGCTCTGTATAAAACAGAAATTG 1063

QY 1598 CTGGCGGAGTAGGCGGTGGTCCACCTGACCCCATGCCAACTCAGAAAGTGA 1653

Db 1064 CTGGCGGAGTAGGCGGTGGTCCACCTGACCCCATGCCAACTCAGAAAGTGA 1119

RESULT 5
US-08-447-430A-40
Sequence 40, Application US/08447430A
Patent No. 5916558
GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Recombinant polypeptides and peptides,

TITLE OF INVENTION: nucleic acids coding for the same and use of these

TITLE OF INVENTION: polypeptides and peptides in the diagnostic of

TITLE OF INVENTION: tuberculosis.

NUMBER OF SEQUENCES: 43

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/447,430A

FILING DATE:

CLASSIFICATION: 424

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 3423 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: plasmid vector

HYPOTHETICAL: NO

US-08-447-430A-40

Query Match 9.2%; Score 152.4; DB 2; Length 3423;
Best Local Similarity 97.6%; Pred. No. 1.2e-30;
Matches 166; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1484 TCTAGAGTCGACCTGCGGATGCAAGCTTGGCTGTTTGGCGGATGAGAGATTTC 1543

Db 363 TCTAGAGTCGACCTGCGGATGCAAGCTTGGCTGTTTGGCGGATGAGAGATTTC 419

QY 1544 AGCCTGATACAGATTAAATCAGAACGCGGCTCTGTATAAAACAGAAATTTCCTGCG 1603

Db 420 AGCCTGATACAGATTAAATCAGAACGCGGCTCTGTATAAAACAGAAATTTCCTGCG 479

QY 1604 GGCAGTAGCGCGGTGGTCCACCTGACCCCATGCCAACTCAGAAAGTGA 1653

Db 480 GGCAGTAGCGCGGTGGTCCACCTGACCCCATGCCAACTCAGAAAGTGA 529

RESULT 6
US-09-485-737B-86
Sequence 86, Application US/09485737B
Patent No. 6350860
GENERAL INFORMATION:

APPLICANT: Buyse, Marie-Ange

APPLICANT: Sablon, Erwin

TITLE OF INVENTION: INTERFERON-gamma-BINDING MOLECULES FOR TREATING SEPTIC SHOCK,

TITLE OF INVENTION: CACHEXIA, IMMUNE DISEASES AND SKIN DISORDERS

FILE REFERENCE: INNS:015

CURRENT APPLICATION NUMBER: US/09/485,737B

CURRENT FILING DATE: 2000-02-14

PRIOR APPLICATION NUMBER: PCT/EP 98/05165

PRIOR FILING DATE: 1998-08-14

PRIOR APPLICATION NUMBER: EPO 98870139.7

PRIOR FILING DATE: 1998-06-18

PRIOR APPLICATION NUMBER: EPO 97870122.5

PRIOR FILING DATE: 1997-08-18

NUMBER OF SEQ ID NOS: 104

SOFTWARE: PatentIn version 3.0

SEQ ID NO 86

LENGTH: 3423

TYPE: DNA

ORGANISM: UNKNOWN

FEATURE:

OTHER INFORMATION: GENOMIC

US-09-485-737B-86

Query Match 9.2%; Score 152.4; DB 4; Length 3423;
Best Local Similarity 97.6%; Pred. No. 1.2e-30;
Matches 166; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1484 TCTAGAGTCGACCTGCGGATGCAAGCTTGGCTGTTTGGCGGATGAGAGATTTC 1543

Db 363 TCTAGAGTCGACCTGCGGATGCAAGCTTGGCTGTTTGGCGGATGAGAGATTTC 419

QY 1544 AGCCTGATACAGATTAAATCAGAACGCGGCTCTGTATAAAACAGAAATTTCCTGCG 1603

Db 420 AGCCTGATACAGATTAAATCAGAACGCGGCTCTGTATAAAACAGAAATTTCCTGCG 479

QY 1604 GGCAGTAGCGCGGTGGTCCACCTGACCCCATGCCAACTCAGAAAGTGA 1653

Db 480 GGCAGTAGCGCGGTGGTCCACCTGACCCCATGCCAACTCAGAAAGTGA 529

RESULT 7
US-08-761-258-7
Sequence 7, Application US/08761258
Patent No. 5756087
GENERAL INFORMATION:

APPLICANT: Ligon, James M.

APPLICANT: Hill, Dwight S.

APPLICANT: Lam, Stephen T.

APPLICANT: Gaffney, Thomas D.

APPLICANT: Torkewitz, Nancy

TITLE OF INVENTION: Genetically Modified Pseudomonas Strains

TITLE OF INVENTION: with Enhanced Biocontrol Activity
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/761,258
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1097 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: pK223-3
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1
OTHER INFORMATION: /note= "BssHI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7
OTHER INFORMATION: /note= "BglII site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 265
OTHER INFORMATION: /note= "EcoRI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 270
OTHER INFORMATION: /note= "SmaI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 275
OTHER INFORMATION: /note= "BamHI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 281
OTHER INFORMATION: /note= "SalI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 287
OTHER INFORMATION: /note= "PstI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 293
OTHER INFORMATION: /note= "XbaI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 299
OTHER INFORMATION: /note= "XhoI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 305
OTHER INFORMATION: /note= "KpnI site"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 311
OTHER INFORMATION: /note= "No. 57560871 site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 319
OTHER INFORMATION: /note= "HindIII site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1086
OTHER INFORMATION: /note= "BglI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1092
OTHER INFORMATION: /note= "BssHI site"
FEATURE:
NAME/KEY: promoter
LOCATION: 185..264
OTHER INFORMATION: /standard_name= "lac"
FEATURE:
NAME/KEY: terminator
LOCATION: 327..752
OTHER INFORMATION: /standard_name= "trnB"
US-08-761-258-7
Query Match 9.2%; Score 152; DB 1; Length 1097;
Best Local Similarity 77.1%; Pred.No. 1e-30;
Matches 185; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1414 TTGTGACGAGTGGACGACAAAGTGTGACTAGTCAAGATGAGAAATTCGAGCTCGGTAC 1473
DB 225 TGTGAAATTGTGAGCGATTAACAATTTCACACAGAAACAAATTCCTCGGGATCCGTCG 284
QY 1474 CCGGGATCCTCTAGAGTGCACCTGCAGCATGCACGCTTGCTGTTTGGCGATGAGA 1533
DB 285 ACCTGACGTGACTGACGAGGTACCGGCGCAAGCTTGCTGTTTGGCGATGAGA 344
QY 1534 GAAGATTTTCAGCTGATACAGATTAAATCAAGACGAGAAACGGTCTGATTAACAGAA 1593
DB 345 GAAGATTTTCAGCTGATACAGATTAAATCAAGACGAGAAACGGTCTGATTAACAGAA 404
QY 1594 TTGGCTGCGCGACATGACGGCGGTGTCCACCTGACCCCATGCCGAATCAGAAAGTAA 1653
DB 405 TTGGCTGCGCGACATGACGGCGGTGTCCACCTGACCCCATGCCGAATCAGAAAGTAA 464
RESULT 8
US-08-977-306-7
Sequence 7, Application US/08977306
Patent No. 5955348
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight S.
APPLICANT: Gaffney, Thomas D.
APPLICANT: Torkewitz, Nancy
APPLICANT: Stafford, Jill M.
TITLE OF INVENTION: Genetically Modified Pseudomonas Strains
TITLE OF INVENTION: with Enhanced Biocontrol Activity
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5955348artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/977,306
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1097 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: pKK223-3
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1
OTHER INFORMATION: /note= "BssHII site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7
OTHER INFORMATION: /note= "BglII site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 265
OTHER INFORMATION: /note= "EcoRI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 270
OTHER INFORMATION: /note= "SmaI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 275
OTHER INFORMATION: /note= "BamHI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 281
OTHER INFORMATION: /note= "Sali site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 287
OTHER INFORMATION: /note= "PstI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 293
OTHER INFORMATION: /note= "XbaI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 299
OTHER INFORMATION: /note= "XhoI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 305
OTHER INFORMATION: /note= "KpnI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 311
OTHER INFORMATION: /note= "No. 5955348I site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 319
OTHER INFORMATION: /note= "HindIII site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1086
OTHER INFORMATION: /note= "BglI site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1092

OTHER INFORMATION: /note= "BssHII site"
FEATURE:
NAME/KEY: promoter
LOCATION: 185..264
OTHER INFORMATION: /standard_name= "tac"
FEATURE:
NAME/KEY: terminator
LOCATION: 327..752
OTHER INFORMATION: /standard_name= "rrnB"
US-08-977-306-7
Query Match 9.2%; Score 152; DB 2; Length 1097;
Best Local Similarity 77.1%; Pred No. 1e-30;
Matches 185; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1414 TTGTGGAGCAGTTGACGCAAGGTGTGACTAGTCAAGATAGGAATTCGAGTCGGTAC 1473
DB 225 TGTGGAATTGTGAGCGGATAACAATTTACACAGGAACAGAAATCCCGGGGATCCGTCG 284
QY 1474 CCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCGACTTGGCTGTTTGGCGGATGAGA 1533
DB 285 ACCTGCAGTCTAGACTCGAGGGTACCGCGCGCAAGCTTGGCTGTTTGGCGGATGAGA 344
QY 1534 GAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAAGCGGTCTGATAAAACAGAA 1593
DB 345 GAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAAGCGGTCTGATAAAACAGAA 404
QY 1594 TTGTGCTGGCGCAGTAGCGCGTGTCCACCTGACCCCATGCCGAACCTCAGAAGTGAA 1653
DB 405 TTGTGCTGGCGCAGTAGCGCGTGTCCACCTGACCCCATGCCGAACCTCAGAAGTGAA 464
RESULT 9
US-08-531-601-3
Sequence 3, Application US/08531601
Patent No. 5688684
GENERAL INFORMATION:
APPLICANT: YOSHIGI, NAOHIRO
APPLICANT: MAEBA, HIDEO
APPLICANT: OKADA, YUKIO
TITLE OF INVENTION: RECOMBINANT BETA-AMYLASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/531,601
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP HEI 6-233086
FILING DATE: 28-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 08/531,601
REFERENCE/DOCKET NUMBER: 2589-031-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6312 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 US-08-531-601-3

Query Match 9.0%; Score 148.8; DB 1; Length 6312;
 Best Local Similarity 98.7%; Pred. No. 1.3e-29;
 Matches 150; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1502 GCATCGAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA 1561
 DB 1747 GCAGCGAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA 1806
 QY 1562 TCAGAACCGAAGAGCGGCTGTGATAAACAAGATTTGCTGGCGGCGATGCGCGGTGCTC 1621
 DB 1807 TCAGAACCGAAGAGCGGCTGTGATAAACAAGATTTGCTGGCGGCGATGCGCGGTGCTC 1866
 QY 1622 CCACCTGACCCCATGCCGAACTCAGAAAGTGAA 1653
 DB 1867 CCACCTGACCCCATGCCGAACTCAGAAAGTGAA 1898

RESULT 10

US-08-859-032-3
 Sequence 3, Application US/08859032
 Patent No. 5863784

GENERAL INFORMATION:
 APPLICANT: YOSHIGI, NAOKIHO
 APPLICANT: MAEDA, HIDEO
 APPLICANT: OKADA, YUKIO
 TITLE OF INVENTION: RECOMBINANT BETA-AMYLASE
 NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
 CITY: ARLINGTON
 STATE: VIRGINIA
 COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/859,032

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/531,601

FILING DATE: 21-SEP-1995

APPLICATION NUMBER: JP HEI 6-233086

FILING DATE: 28-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F

REGISTRATION NUMBER: 08/531,601

REFERENCE/DOCKET NUMBER: 2589-031-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 6312 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

US-08-859-032-3

Query Match 9.0%; Score 148.8; DB 2; Length 6312;
 Best Local Similarity 98.7%; Pred. No. 1.3e-29;
 Matches 150; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1502 GCATCGAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA 1561
 DB 1747 GCAGCGAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA 1806
 QY 1562 TCAGAACCGAAGAGCGGCTGTGATAAACAAGATTTGCTGGCGGCGATGCGCGGTGCTC 1621
 DB 1807 TCAGAACCGAAGAGCGGCTGTGATAAACAAGATTTGCTGGCGGCGATGCGCGGTGCTC 1866
 QY 1622 CCACCTGACCCCATGCCGAACTCAGAAAGTGAA 1653
 DB 1867 CCACCTGACCCCATGCCGAACTCAGAAAGTGAA 1898

RESULT 11

US-09-242-901-3
 Sequence 3, Application US/09242901
 Patent No. 6322783

GENERAL INFORMATION:

APPLICANT: TAKAHASHI, SEISHI

TITLE OF INVENTION: NOVEL BACTERIOPHAGES, METHOD FOR SCREENING SAME AND

TITLE OF INVENTION: BACTERICIDAL COMPOSITIONS USING SAME, AND DETECTION

TITLE OF INVENTION: KITS USING SAME

FILE REFERENCE: 3000-0001

CURRENT APPLICATION NUMBER: US/09/242,901

CURRENT FILING DATE: 1998-02-26

EARLIER APPLICATION NUMBER: PCT/JP97/02957

EARLIER FILING DATE: 1997-08-26

EARLIER APPLICATION NUMBER: 8/261132

EARLIER FILING DATE: 1996-08-26

EARLIER APPLICATION NUMBER: 9/135716

EARLIER FILING DATE: 1997-04-19

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 3

LENGTH: 518

TYPE: DNA

ORGANISM: Bacteriophage

FEATURE:

OTHER INFORMATION: "n" at various positions throughout the sequence may be

OTHER INFORMATION: a, t, g, c other or unknown

US-09-242-901-3

Query Match 8.9%; Score 147.2; DB 4; Length 518;
 Best Local Similarity 98.0%; Pred. No. 1.4e-29;
 Matches 149; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1502 GCATCGAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA 1561
 DB 189 GAAAGAAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAA 248
 QY 1562 TCAGAACCGAAGAGCGGCTGTGATAAACAAGATTTGCTGGCGGCGATGCGCGGTGCTC 1621
 DB 249 TCAGAACCGAAGAGCGGCTGTGATAAACAAGATTTGCTGGCGGCGATGCGCGGTGCTC 308
 QY 1622 CCACCTGACCCCATGCCGAACTCAGAAAGTGAA 1653
 DB 309 CCACCTGACCCCATGCCGAACTCAGAAAGTGAA 340

RESULT 12

US-08-447-430A-41
 Sequence 41, Application US/08447430A
 Patent No. 5916558

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Recombinant polypeptides and peptides,

TITLE OF INVENTION: nucleic acids coding for the same and use of these

TITLE OF INVENTION: polypeptides and peptides in the diagnosis of

TITLE OF INVENTION: tuberculosis.

NUMBER OF SEQUENCES: 43

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,430A
; FILING DATE:
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: plasmid vector
; HYPOTHETICAL: NO
; US-08-447-430A-41

Query Match      8.9%; Score 147; DB 2; Length 3474;
Best Local Similarity 100.0%; Pred. No. 3.1e-29;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1507 CAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGA 1566
Db 436 CAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGA 495
QY 1567 ACGCAGAAGCGGTCTGATAAAACAGAAATTTGCTGGCGGAGTAGCGCGTGTCCACC 1626
Db 496 ACGCAGAAGCGGTCTGATAAAACAGAAATTTGCTGGCGGAGTAGCGCGTGTCCACC 555
QY 1627 TGACCCCATGCCGAATCTCAGAAAGTAA 1653
Db 556 TGACCCCATGCCGAATCTCAGAAAGTAA 582

RESULT 13
US-08-318-837-10
; Sequence 10, Application US/08318837
; Patent No. 5981277
; GENERAL INFORMATION:
; APPLICANT: FRANSEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,
; APPLICANT: ANDRE; VAN HEUVERSWYN, HUGO
; TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID
; TITLE OF INVENTION: CODING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR THERAPY
; TITLE OF INVENTION: IMMUNOLOGY
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,837
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP 93/01022
; FILING DATE: 28-APR-1993
; APPLICATION NUMBER: 92.401.231.3
; FILING DATE: 30-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000

```

```

; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3474 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: pMTNP-MPH
; US-08-318-837-10

Query Match      8.9%; Score 147; DB 2; Length 3474;
Best Local Similarity 100.0%; Pred. No. 3.1e-29;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1507 CAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGA 1566
Db 436 CAAGCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGA 495
QY 1567 ACGCAGAAGCGGTCTGATAAAACAGAAATTTGCTGGCGGAGTAGCGCGTGTCCACC 1626
Db 496 ACGCAGAAGCGGTCTGATAAAACAGAAATTTGCTGGCGGAGTAGCGCGTGTCCACC 555
QY 1627 TGACCCCATGCCGAATCTCAGAAAGTAA 1653
Db 556 TGACCCCATGCCGAATCTCAGAAAGTAA 582

RESULT 14
US-08-122-458D-8
; Sequence 8, Application US/08122458D
; Patent No. 6387372
; GENERAL INFORMATION:
; APPLICANT: Cocito, Carlo
; APPLICANT: Coene, Marc
; APPLICANT: DeKessel, Myrian
; APPLICANT: Gilot, Philippe
; TITLE OF INVENTION: Polypeptides from Mycobacterium
; TITLE OF INVENTION: Paratuberculosis
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 6387372west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/122,458D
; FILING DATE: 24-SEP-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP92/00661
; FILING DATE: 24-MAR-1992
; APPLICATION NUMBER:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91400798.4
; FILING DATE: 25-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.111USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 8:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 3474 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-122-458D-8

Query Match 8.9%; Score 147; DB 4; Length 3474;
Best Local Similarity 100.0%; Pred.No. 3.1e-29;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1507 CAAGCTTGCTGCTTTGGGGATGAGAGAAATTTCAAGCTGATACAGATTAAATCAGA 1566
DB 436 CAAGCTTGCTGCTTTGGGGATGAGAGAAATTTCAAGCTGATACAGATTAAATCAGA 495
QY 1567 ACGCAGAACGGGTCTGATTAACAGAAATTTGCTGGCGGAGTAGCGCGGTGTCACC 1626
DB 496 ACGCAGAACGGGTCTGATTAACAGAAATTTGCTGGCGGAGTAGCGCGGTGTCACC 555
QY 1627 TGACCCCATGCCGAATCAGAAAGTGAA 1653
DB 556 TGACCCCATGCCGAATCAGAAAGTGAA 582

RESULT 15
US-09-131-028A-6/c
Sequence 6, Application US/09131028A
Patent No. 6287866
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Lemmel, Steven A.
APPLICANT: Leonard, Amanda Eun-Yeong
APPLICANT: Chaudhary, Sunita
TITLE OF INVENTION: BETA-CASEIN EXPRESSING CONSTRUCTS
FILE REFERENCE: 6004.US.P1
CURRENT APPLICATION NUMBER: US/09/131,028A
CURRENT FILING DATE: 1998-08-07
PRIOR APPLICATION NUMBER: US 08/064,440
PRIOR FILING DATE: 1993-05-21
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 8430
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (6124)...(6939)
OTHER INFORMATION: KANR - kanamycin resistance in forward orientation
US-09-131-028A-6

Query Match 8.9%; Score 147; DB 4; Length 8430;
Best Local Similarity 100.0%; Pred.No. 4.1e-29;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1507 CAAGCTTGCTGCTTTGGGGATGAGAGAAATTTCAAGCTGATACAGATTAAATCAGA 1566
DB 7428 CAAGCTTGCTGCTTTGGGGATGAGAGAAATTTCAAGCTGATACAGATTAAATCAGA 7369
QY 1567 ACGCAGAACGGGTCTGATTAACAGAAATTTGCTGGCGGAGTAGCGCGGTGTCACC 1626
DB 7368 ACGCAGAACGGGTCTGATTAACAGAAATTTGCTGGCGGAGTAGCGCGGTGTCACC 7309
QY 1627 TGACCCCATGCCGAATCAGAAAGTGAA 1653
DB 7308 TGACCCCATGCCGAATCAGAAAGTGAA 7282

Search completed: January 16, 2003, 08:53:08
Job time : 88 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 08:18:51 ; Search time 70 Seconds
(without alignments)
10528.861 Million cell updates/sec

Title: US-09-830-111A-1
Perfect score: 1653
Sequence: 1 ttttggggtcgaaagtct.....atgccgaactcagaagtga 1653

Scoring table: IDENTITY_NUC
Gapop 10.0 , Capext 1.0

Searched: 393868 seqs, 222934149 residues

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMP.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMP.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCT05_PUBCOMP.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMP.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMP.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMP.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMP.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	179.2	10.8	5926	9	US-10-024-809-3
C 2	179.2	10.8	5926	9	US-10-024-809-4
C 3	145.4	8.8	6271	10	US-09-993-292A-1
C 4	145.4	8.8	8908	10	US-09-993-292A-18
C 5	138	8.3	425	10	US-09-960-428-10
C 6	136	8.2	1173	10	US-09-887-576-678
C 7	136	8.2	1179	9	US-09-842A-990
C 8	130.8	7.9	7416	10	US-09-900-575-46
C 9	120.8	7.3	894	9	US-09-938-842A-1124
C 10	115.4	7.0	799	10	US-09-912-020-27
C 11	79.2	4.8	927	10	US-09-974-300-999
C 12	68.2	4.1	972	10	US-09-815-242-9636
C 13	64.2	3.9	787	12	US-10-040-037-1
C 14	63	3.8	1323	12	US-10-013-036-3
C 15	61.6	3.7	103	10	US-09-179-536B-267
C 16	61.6	3.7	513	10	US-09-974-300-5444
C 17	61.6	3.7	1361	10	US-09-393-634-14
C 18	61.4	3.7	293	9	US-10-132-561-11
C 19	61.4	3.7	293	9	US-10-132-561-12

20	61.4	3.7	458	9	US-10-132-561-14	Sequence 14, Appl
C 21	61.4	3.7	458	9	US-10-132-561-14	Sequence 14, Appl
C 22	61.4	3.7	459	9	US-10-132-561-13	Sequence 13, Appl
C 23	61.4	3.7	600	9	US-10-132-561-5	Sequence 5, Appl
C 24	61.4	3.7	789	9	US-10-132-561-16	Sequence 16, Appl
C 25	61.4	3.7	789	9	US-10-132-561-16	Sequence 16, Appl
C 26	61.4	3.7	790	9	US-10-132-561-15	Sequence 15, Appl
C 27	61.4	3.7	790	9	US-10-132-561-15	Sequence 15, Appl
C 28	61.4	3.7	2686	10	US-09-785-269-7	Sequence 7, Appl
C 29	61.4	3.7	71	10	US-09-380-932-3	Sequence 3, Appl
C 30	61	3.7	206	9	US-10-033-297-27	Sequence 27, Appl
C 31	61	3.7	206	9	US-10-081-806-32	Sequence 32, Appl
C 32	61	3.7	586	10	US-09-764-847-1873	Sequence 1873, Ap
C 33	61	3.7	972	10	US-09-815-242-6224	Sequence 6224, Ap
C 34	61	3.7	6741	10	US-09-794-384A-14	Sequence 14, Appl
C 35	60.8	3.7	560	10	US-09-772-134B-26	Sequence 26, Appl
C 36	60	3.6	75	10	US-09-955-649-6	Sequence 6, Appl
C 37	60	3.6	75	10	US-09-973-013-6	Sequence 6, Appl
C 38	60	3.6	119	8	US-08-852-020-15	Sequence 15, Appl
C 39	60	3.6	149	12	US-10-012-211A-5	Sequence 5, Appl
C 40	60	3.6	298	9	US-09-808-124A-4	Sequence 4, Appl
C 41	60	3.6	298	10	US-09-845-157-6	Sequence 6, Appl
C 42	60	3.6	384	10	US-09-757-992-6	Sequence 6, Appl
C 43	60	3.6	384	10	US-09-757-992-13	Sequence 13, Appl
C 44	60	3.6	910	8	US-08-852-020-5	Sequence 5, Appl
C 45	59.6	3.6	102	9	US-10-015-219-106	Sequence 106, App

ALIGNMENTS

RESULT 1
US-10-024-809-3/c
; Sequence 3, Application US/10024809
; Publication No. US20020188105A1
; GENERAL INFORMATION:
; APPLICANT: CRAIG, NANCY L
; TITLE OF INVENTION: GAIN OF FUNCTION MUTATIONS IN
; ATP-DEPENDENT TRANSDUCTION PROTEINS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Anne Brown (Alston & Bird, LLP)

STREET: 3605 Glenwood Ave.

CITY: Raleigh

STATE: NC

COUNTRY: USA

ZIP: 27608

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/10/024,809

FILING DATE: 19-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/027,169

FILING DATE: 1998-FEB-20

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Anne

REGISTRATION NUMBER: 36,463

REFERENCE/DOCKET NUMBER: 5789-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919 420 2205

TELEFAX: 919 861 3175

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 5926 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "pew delta R. adj to 1"
SEQUENCE DESCRIPTION: SEQ ID NO: 3
US-10-024-809-3

Query Match 10.8%; Score 179.2; DB 9; Length 5926;
Best Local Similarity 91.3%; Pred. No. 6.5e-45;
Matches 190; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1446 GGTCAAGATAGGAATTGAGCTCGGTACCCGGGATCCTTGAAGTCGACCTGCAGGCAT 1505
DB 4282 GTTATTAAAGCTGTCGCCAGCAGACTTTCTTGGGATCCTCTAGAGTCGACCTGCAGGCAT 4223
QY 1506 GCAAGCTTGGCTGTTTGGCCGATGAGAGAATTTTGAAGCTTGAATTAATCAG 1565
DB 4222 GCAAGCTTGGCTGTTTGGCCGATGAGAGAATTTTGAAGCTTGAATTAATCAG 4163
QY 1566 AACGAGAAGCGGCTGTGATTAACAAGATTTTGCCTGGCGGAGTAGCGGCTGTCCAC 1625
DB 4162 AACGAGAAGCGGCTGTGATTAACAAGATTTTGCCTGGCGGAGTAGCGGCTGTCCAC 4103
QY 1626 CTGACCCCATGCCGAAGTCAAGTCAA 1653
DB 4102 CTGACCCCATGCCGAAGTCAAGTCAA 4075

RESULT 2

US-10-024-809-4/c
Sequence 4, Application US/10024809
Publication No. US20020188105A1

GENERAL INFORMATION:

APPLICANT: CRAIG, NANCY L
TITLE OF INVENTION: GAIN OF FUNCTION MUTATIONS IN
ATP-DEPENDENT TRANSDUCTION PROTEINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anne Brown (Alston & Bird, LLP)
STREET: 3605 Glenwood Ave.
CITY: Raleigh
STATE: NC
COUNTRY: USA
ZIP: 27608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/024,809
FILING DATE: 19-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/027,169
FILING DATE: 1998-FEB-20
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Anne
REGISTRATION NUMBER: 36,463
REFERENCE/DOCKET NUMBER: 5789-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2205
TELEFAX: 919 881 3175

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 5926 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "pew-delta"
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-024-809-4

Query Match 10.8%; Score 179.2; DB 9; Length 5926;
Best Local Similarity 91.3%; Pred. No. 6.5e-45;

Matches 190; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1446 GGTCAAGATAGGAATTGAGCTCGGTACCCGGGATCCTTGAAGTCGACCTGCAGGCAT 1505
DB 3918 GTTATTAAAGCTGTCGCCAGCAGACTTTCTTGGGATCCTCTAGAGTCGACCTGCAGGCAT 3859
QY 1506 GCAAGCTTGGCTGTTTGGCCGATGAGAGAATTTTGAAGCTTGAATTAATCAG 1565
DB 3858 GCAAGCTTGGCTGTTTGGCCGATGAGAGAATTTTGAAGCTTGAATTAATCAG 3799
QY 1566 AACGAGAAGCGGCTGTGATTAACAAGATTTTGCCTGGCGGAGTAGCGGCTGTCCAC 1625
DB 3798 AACGAGAAGCGGCTGTGATTAACAAGATTTTGCCTGGCGGAGTAGCGGCTGTCCAC 3739
QY 1626 CTGACCCCATGCCGAAGTCAAGTCAA 1653
DB 3738 CTGACCCCATGCCGAAGTCAAGTCAA 3711

RESULT 3

US-09-993-292A-1/c
Sequence 1, Application US/09993292A
Patent No. US20020146430A1

GENERAL INFORMATION:

APPLICANT: James E. Galen
TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
PROTEINS
FILE REFERENCE: UOFMD.007A
CURRENT APPLICATION NUMBER: US/09/993,292A
CURRENT FILING DATE: 2001-11-23
PRIOR APPLICATION NUMBER: 60/252,516
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 6271
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: pSEC84 Expression Plasmid
US-09-993-292A-1

Query Match 8.8%; Score 145.4; DB 10; Length 6271;
Best Local Similarity 99.3%; Pred. No. 2e-34;
Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1507 CAAGCTTGGCTGTTTGGCCGATGAGAGAATTTTCAAGCTTGAATTAATCAG 1566
DB 2571 CCAGCTTGGCTGTTTGGCCGATGAGAGAATTTTCAAGCTTGAATTAATCAG 2512
QY 1567 ACGCAGAAGCGGCTGTGATTAACAAGATTTTGCCTGGCGGAGTAGCGGCTGTCCAC 1626
DB 2511 ACGCAGAAGCGGCTGTGATTAACAAGATTTTGCCTGGCGGAGTAGCGGCTGTCCAC 2452
QY 1627 TGACCCCATGCCGAAGTCAAGTCAA 1653
DB 2451 TGACCCCATGCCGAAGTCAAGTCAA 2425

RESULT 4

US-09-993-292A-18/c
Sequence 18, Application US/09993292A
Patent No. US20020146430A1

GENERAL INFORMATION:

APPLICANT: James E. Galen
TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
PROTEINS
FILE REFERENCE: UOFMD.007A
CURRENT APPLICATION NUMBER: US/09/993,292A
CURRENT FILING DATE: 2001-11-23
PRIOR APPLICATION NUMBER: 60/252,516


```

? APPLICANT: Wang, X.
? APPLICANT: Cooper, Bret
? TITLE OF INVENTION: Promoters for regulation of plant expression
? FILE REFERENCE: 1360.001US1
? CURRENT APPLICATION NUMBER: US/09/887,576
? CURRENT FILING DATE: 2001-06-25
? PRIOR APPLICATION NUMBER: US 60/213,848
? PRIOR FILING DATE: 2000-06-23
? PRIOR APPLICATION NUMBER: US 60/214,087
? PRIOR FILING DATE: 2000-06-23

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; SEQ ID NO 678
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-678

Query Match      8.2%; Score 136; DB 10; Length 1173;
Best Local Similarity 50.3%; Pred. No. 6.2e-32;
Matches 380; Conservative 0; Mismatches 355; Indels 21; Gaps 1;

Qy 671 CAACGGCGGTTGGCTGAGATCACCGAGATGCATCATCCTCCACGAGC 730
    ||| ||||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy	731	GTATTGACAGCTTCCGAGACCGACGAGAAAACGACCATCCGGAAACGAGGCATTCGGAAAC	790
Db	469	GTGTTAGACAGAGTGATGCTCGAAGAGGAAAGAAAACAGTTCTATGAGCTTTTCGGCACA	528
Qy	791	AAGATGCCGATTTTGGCTGTGTGATTCTTGTTGGGACGGCGCTGTGTTGCATTGCCGAGG	850
Db	529	AGAGTAGCGGTGTAGCTGGAGATTTTCATGTTTGCCTCAAGCGTCATCGTACTTAGCAAAAT	588

Qy	851	TTGCGCAATCGGAGGTGATTGAGCTTTTGCTACTGTTATTGCAAACTTGGTTGAGGA	910
Db	589	CTCAGAATCTTGAAGTTATTAAGCTCATCAGTCAGGTGATCAAGAGCTTTGCAAGCGGA	648
Qy	911	GAGTTTCATGCAGTTTGA AAAATAC TGTGTTGATGATCGGATTGAGGCTACGGCGACGCAGGA	970
Db	649	GAGATAAGCAGGCGTCCAGCTTATTTGACTGCG-----ACACC	687

Db 688 AAGCTCGACGAGTACTACTCAAAAGTTTCTACAAGACAGCCTCTTTAGTGGCTGGAGC 747

Ov 1031 TGCAGAGCAAGTCGGCTTCTGGGTGGTGCTACGCCGTGAGGTTGCTGATGCTGCTTATGCT 1090

Db	748	ACCAAGAGAGTGCCATTTTTCAGCAGAGTTGAGCCTGATGTGACAGAAACAATATCCGAG	807
Qy	1091	TACGGAAAGGAACCTTGTTTGGCATTCAGATCTCGACACATGCTCGACTACACCGTC	1150
Db	808	TTTGGGAAGAATCTCGCTCTCTTTCCAGATAGTTGATGATATTTTGGATTTTCACTCAG	867
Qy	1151	TCGCGTACCGACCTCTCGCTBAGCCCGCGCGTGCACCTCCAGCTCGCTCGCCACCGCG	1210

D_b 868 TCACAGAGCAGCTCGGGAAGCCAGCAGGGAGTGAATTTGGCTAAGGTAACTTAACAGCA 927
D_c 1211 CCCCCCCCCTTTTCGCATGCCAACCCNCCNCCGCGCATCGGCTCCGATCTGTCACGCCCAACTTC 1270

Db	928	CCTGTGATTTTCGCTCTGGAGAGGGCCAAAGGCTATAGAGAGATCATTTGAGTCAGAGTTT	987
Qy	1271	TCTGACCCAGGAGACGTCGAGCGTGCACGGCAGTTGGTCGAGAAAAGTGATGGATTTGGAG	1330
Db	988	TGTGAGCGGGTCTCTCGAAGAGCCGATTGAACGGTGACAAAAGTGGGGGATTAAAG	1047
Qy	1331	AAGACGAGAGCCTTGGCGGAGGAGTATGCCAGAAAGCGTTTGGATGCAATTCGGACGTTTC	1390
Db	1048	AGAGCAAGAATTTGGCTTAGGGAGAAAGCTGATGACGCTATAAAGAAATCTACAGTGCTA	1107
Qy	1391	CCGAGAGATCCGCGACCGAAGGCTTTCCGAGCAGTTG	1426

Db 1108 CCTCGAAGTGGCTTCAGAGTGGCTCTAGAAAGATATG 1143

RESULT 7

US-09-938-842A-990
 ; Sequence 990, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; FILE REFERENCE: SCRIPT300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 990
 ; LENGTH: 1179
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-938-842A-990

Query Match 8.2%; Score 136; DB 9; Length 1179;
 Best Local Similarity 50.3%; Pred. No. 6.2e-32;
 Matches 380; Conservative 0; Mismatches 355; Indels 21; Gaps 1;

Qy 671 CAACGCGGCTTGGCTGAGATCAAGAGATGATCCATGACATCACTCTCCACGACGAC 730
 Db 415 CACCGCGCTTGGCTGAGATCAATCGAGATACACATCGAAGCTTGATACACGCGAT 474
 Qy 731 GTTATCGACGCTTCCGAGACGACGAAACGACCATCCGAAACCGAGCATTTGGAAAC 790
 Db 475 GTTATGAGAGAGTACATCGGAAAGAAAGAAACGTTCAATGAGCTTTTCGACACA 534
 Qy 791 AAGATGCGCATTTTGGCTGAGATTTCTTTGGGACGCGGCGCTCTGTTGCAATTGGCGAG 850
 Db 535 AGAGTAGCGGCTGCTAGCTGAGATTTCACTTTGCTCAAGCGCATGTAAGTAAACAAT 594
 Qy 851 TTGCGCAATCCGAGAGTATGAGCTTTGGCTACTGTTATTCGAAACTTTGGTTAGGGA 910
 Db 595 CTCGGAATCTTGAAGATTATTAAGCTCATCACTGATGATCAAACTTTGCAAGCGGA 654
 Qy 911 GAGTTCATGCACTTGAAGAAATCTGTTGATGATGCGATTGAGGCTCGGCGACGACGAA 970
 Db 655 GAGATTAAGCAGGCTGCCAGCTTATTGACTGCG-----ACACC 693
 Qy 971 ACGTTGATTACTATTGGCAGAGACTTACTTAAGACTGCGTCTTGAATTGCCAAGTGC 1030
 Db 694 AAGCTCGAGAGTACTTACTCAAAAGTTTCTTAAGACAGCGCTTTAGTGGCTGCGAGC 753
 Qy 1031 TGCAGAGCAAGTCCGCTTGGGTGCTGACGCTGAGAGTTCGATGCTGTTATGCT 1090
 Db 754 ACCAAGAGGCTGCCATTTTTCAGCAGAGTTGAGCCTGATGTCAGAACAAATGTGCGAG 813
 Qy 1091 TACGGAAGAACCTTGGTTGGATTCGATCGATCGTGAAGCAATGCTGACTACACCGTC 1150
 Db 814 TTGGGAAGAAATCTGCTCTCTTTCAGATAGTTGATGATTTTGGATTTCACTCAG 873
 Qy 1151 TCCGCTACCGACCTCGGTAAAGCCGCGGTGACAGCTTCAGCTCGTCTCGCAACCGCG 1210
 Db 874 TCGAAGAGACACTCGGGAAGCAGAGGAGTGAATTTGGCTAAAGTAACTTAACAGA 933
 Qy 1211 CCGGCTCTTTCGATGGAAGACCAAGCGGAGCTCGGTCCCATGATCAAGCGCAAGTTTC 1270

Db 934 CCTGATTTTTCGCTCTGAGAGAGGAGCCAGGCTTAAGAGATCATTTGATCAGACTTT 993

Qy 1271 TCTGACCCAGAGAGAGCTGACGCTGACGCGAGTTGTTGAGAAAAAGTATGATTTGAG 1330

Db 994 TGTGAGCGGCTTCTCTGGAAGAACGATTGAGCGGTGACAAAAGGTGGGGGATTAAAG 1053

Qy 1331 AAGACGAGAGCTTGGCCGAGAGATATCCAGAAAGGCTTGGATGATTCGACGCTTC 1390

Db 1054 AGAGCAAGAAATTTGGCTTGAAGGAGAAAGCTGATGACGCTATAAGAAATCTACAGTCTCA 1113

Qy 1391 CCGAGAGTCCGACGAGAGGCTTTGGAGCAGTTG 1426

Db 1114 CCTGAAAGTGGCTTCAGAGTGGCTCTCTAGAAAGATATG 1149

RESULT 8

US-09-900-575-46
 ; Sequence 46, Application US/09900575
 ; Patent No. US20020150587A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Langermann, Solomon
 ; APPLICANT: Revel, Andrew
 ; APPLICANT: Auguste, Christine
 ; APPLICANT: Burlein, Jeanne
 ; TITLE OF INVENTION: Film Adhesin Proteins and Methods of Use
 ; FILE REFERENCE: 469201-549
 ; CURRENT APPLICATION NUMBER: US/09/900,575
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: US/60/216,750
 ; PRIOR FILING DATE: 2000-07-07
 ; NUMBER OF SEQ ID NOS: 64
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 46
 ; LENGTH: 7416
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Sequence of Plasmid pCGA126-1
 ; US-09-900-575-46

Query Match 7.9%; Score 130.8; DB 10; Length 7416;
 Best Local Similarity 98.5%; Pred. No. 7.5e-30;
 Matches 132; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1512 TTGCGTGTGTTGGCGAGTGAAGAAATTTCAAGCTGATACAGATTAATCAGAACGCA 1571
 Db 2255 TTCCCTGTTTGGCGAGTGAAGAAATTTCAAGCTGATACAGATTAATCAGAACGCA 2314
 Qy 1572 GAAGCGTCTGATAAACAGAAATTTGCTGCGCGAGTAGCGGCTGCTCCCACTGACC 1631
 Db 2315 GAAGCGTCTGATTAACAGAAATTTGCTGCGCGAGTAGCGGCTGCTCCCACTGACC 2374
 Qy 1632 CCATGCCGAACTCA 1645
 Db 2375 CCATGCCGAACTCA 2388

RESULT 9

US-09-938-842A-1124
 ; Sequence 1124, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; FILE REFERENCE: SCRIPT300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24

```

; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1124
; LENGTH: 894
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1124

Query Match
Best Local Similarity 7.3%; Score 120.8; DB 9; Length 894;
Matches 242; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 954 CTACGGCAGCGAGAAACGTTTCGATTACTTATTTGAGAGACTTACTTTGAAGACTGCGT 1013
DB 374 CAACCGAGCAGCGTTATAGTATGGACTACTTACATGCAGAAGACATATTATAAGACAGCAT 433

QY 1014 CTTGATTGCCAAGTCGTCAGAGCAAGTGCCTTCTGGGTGTGCTACGCTGAGGTTG 1073
DB 434 CGCTAATCTTAACAGCTGCAAGCTGTTGCCGTCTCTCACTGGACAAACAGCAGAAAGTTG 493

QY 1074 CTGATCTGCTTATGCTTACGGAAGAACCTTGTTTGGCATTCAGATCGTCGACGACA 1133
DB 494 CGGTGTTAGCTTTTGGTATGGAGAACTTGGGTTTAGCAITTCCTAATTAATAGACGACA 553

QY 1134 TGCTGACTACACGCTCTCGCTTACCGACCTCGGTAAAGCCGCGGTGCAGACCTCCAGC 1193
DB 554 TTCTTGATTTCAGGGCAGCATCTGCTCTCTCGGAAGGATGTTGTTCAGATATTCGCC 613

QY 1194 TGGTCTCGCCACCGCGGCGCTCTTCGATGGAAGACACACGCGGAGCTCGGTCCCA 1253
DB 614 ATGGAGTCAATACAGAGCCCAATCTCTTTGCCATGGAAGAGTTTCTCACTACGCGAAG 673

QY 1254 TGATCAAGCGCAAGTTCTCTGACCCAGGAGACGTCGCGGTGCACGCGAGTTGGTCGAGA 1313
DB 674 TTGTTGATCAAGTTGAAAAGATCTTAGGAATGTTGACATGCTTTAGAGATATCTTTGGGA 733

QY 1314 AAAGTGATGGATTGGAGAGACGAGAGCCTTGGCGGAGGAGTATGCCAGAGGCGGTTGG 1373
DB 734 AGAGCAAGGGAATACAGAGGCAAGAGAAATAGCCATGGAACATGCAATCTAGCAGCAG 793

QY 1374 ATGCAATTCGAGGTTCCCGGAGA 1397
DB 794 CTGCAATCGGGTCTCTACCTGAAA 817

RESULT 10
US-09-912-020-27
; Sequence 27, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27

; LENGTH: 799
; TYPE: DNA
; ORGANISM: E. Coli
; US-09-912-020-27

Query Match
Best Local Similarity 7.0%; Score 115.4; DB 10; Length 799;
Matches 116; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1537 GATTTTCAGCCTGATACAGATTAAATTCAGAACGCGAAGCGGTCTCTGATAAACAGAAATTT 1596
DB 475 GCTTTTCAGCCTGATACAGATTAAATTCAGAACGCGAAGCGGTCTCTGATAAACAGAAATTT 534

QY 1597 GCTTGGCGCAGTAGCGCGGTGTCCTCCACTGACCCCATGCCGAACTCAGAAAGTGAA 1653
DB 535 GCTTGGCGCAGTAGCGCGGTGTCCTCCACTGACCCCATGCCGAACTCAGAAAGTGAA 591

RESULT 11
US-09-974-300-999
; Sequence 999, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 999
; LENGTH: 927
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; US-09-974-300-999

Query Match
Best Local Similarity 4.8%; Score 79.2; DB 10; Length 927;
Matches 341; Conservative 0; Mismatches 363; Indels 24; Gaps 2;

QY 694 GGAGATGATCCATCGCAGCATCACTCTCCACGACGAGTTATCGCGCTTCCGAGACCAG 753
DB 221 GGAATTGATACATATGGCTTCGTCTGTACACGACGATGTCATCGATGCGGAGCTGAG 280

QY 754 ACGAAACGCCATCCGAAACACGAGCATTCGGAACCAAGATGGCGGATTTTGGCTGGTGA 813
DB 281 AGAGGACAGCGCACAATCAAGCGAAGTGGATTAACCGGATTCGAATGTATACAGGTGA 340

QY 814 TTTCTTTTGGGACCGGCGCTCTGTTGATGTCGAGGTTTCCGCAATCCGAGAGTGATGA 873
DB 341 TTATTGTTGCTCGCTCACTTGAAGCGGATGACAAAAATAAACGAAACCGAACCCCATGA 400

QY 874 GCTTTTGGCTACTGTTATTGCAAACTTGTGTTGAGGAGAGTTTCATGCGAGTTGAAAATAC 933
DB 401 AATCTTGTCCAAGGCGGATTTGCGAAGTCTGTCTTGGAGAAATTTGAACAAATTAAGATAA 460

QY 934 TGTGATGATGCGATTGAGGCTACGCGCAGCAGGAAACGTTTCGATTACTATTTCAGAA 993
DB 461 ATATAAT-----ATGAGCAAAATCTCAGAACCTATTTTGAGAAG 499

QY 994 GACTTACTTGAAGACTGCGTCTTGAATGCCAAGTCGTGCAGAGCAAGTCCGCTTCTGGG 1053
DB 500 AATCAGGCGCAAGACAGACACTTCTAATCGCTGAGCTGCCAGCTTGGAGCCATTCCTGC 559

QY 1054 TGGTGCTACGCTGAGTTGCTGATGCTCTTATGCTTACGGAAGCAACTTGGTTTGGC 1113
DB 560 TGGTGCGGTTGAAAACAACTTCATAAAAAGTGTGTAAGTCTGTTTGGGTATTTATGTCGGCATGTC 619

```

QY 1114 ATTCAGATCGTCAGCAGCATCTGACTACACCGCTCTCCGCTACCGACCTGGTAAGCC 1173
 DB 620 CTATCAATTTATTTGATGACATCTCGATTTTCATCAACGAGAAAAAGCTTGGCAAGCA 679
 QY 1174 CGCCGCTGACAGCTCCAGCTGCTGCTGCCACCGCGCCGCTCTTGGCATGGAAGCA 1233
 DB 680 TGTGGGACCGCATCTGCTTCAAGGGAGATGTCACGCTTCCGATTTATGTCACGGAAATC 739
 QY 1224 CCACGCCGAGCTCGGCTCCCATGATCAACGCGCAA---GTTCTCTGACCCAGAGAGCTCGA 1290
 DB 740 GCTTAGTAAAAAGCGCGCTGATGTTGGTCAACAGTAGAAACCATCTCGAAGACATGCA 799
 QY 1291 GCGTCAACGCGAGTGTGTCGAGAAAAAGTAGATGAGGAAGACGAGCCTTGGCGGA 1350
 DB 800 GCGAGTGAATCGAAGAGCTGAGAAAAAGATGTCATTTGAACATCGTTCAGGGGTCACTGA 859
 QY 1351 GGAGTATGCCGGAAGGCTTGGATGCAATTCGACGTTCCCGAGAGATCCCGGACCGGA 1410
 DB 860 AATGATATTMAACAAGGCTTGCATTTGGTGAAACGAAACCGATMAACCGGAGCGCCAC 919
 QY 1411 GCGTTGG 1418
 DB 920 CCGAGTGG 927

RESULT 12

US-09-815-242-9636
 Sequence 9636, Application US/09815242
 Patent No. US2002061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 9636
 LENGTH: 972
 TYPE: DNA
 ORGANISM: Salmonella typhi
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(972)
 US-09-815-242-9636

Query Match 4.1%; Score 68.2; DB 10; Length 972;
 Best Local Similarity 45.8%; Pred. No. 5.9e-11;
 Matches 348; Conservative 0; Mismatches 378; Indels 33; Gaps 2;

QY 667 CTCGCAACGGCGGTTGGCTGATCAGAGATGATTCATGACAGATCACTCTCCACGA 726
 DB 192 CGCGACGTCATCTATCGCGCTCTGATGAGTTTATCCACGCGGACGCTGCTACATGA 251
 QY 727 CGACGTTATCGACGCTTCCGAGACGAGAAACGACCATCCGAAAAACGAGGATTCGG 786
 DB 252 TGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 311
 QY 787 AAACAAGATGGGATTTTGGCTGATGATTTCTTGTGGAGAGGCGCTGTGTGATGGC 846
 DB 312 TAATGCCCGCCAGCGTGTGTGCGGAGCTTATATATACCCGCGCTTCCAGATGATGAC 371
 QY 847 GAGGTTGGCAATCCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 906
 DB 372 CAGCTCGGCTGTTAAAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATG 431
 QY 907 GGGAGATTCAATGACGTTGAAAAATCTGTTGATGATGATGATGATGATGATGATGATGATG 966
 DB 432 AGCGAAGTGTGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 474
 QY 967 GGAACGTTGATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026
 DB 475 ---ACGAAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 530
 QY 1027 GTGCTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
 DB 531 CGCGCGCCAGTGTCTCGGATTTCTTCCGTTGTACGCTTGACGAGAAAGGTTGCA 590
 QY 1087 TGCTTACGGAAGAACTTGTGTTTGCATTCAGATTCGTCAGCATGCTGACTGACTGAC 1146
 DB 591 GGACTATGCGCGCTTACGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 650
 QY 1147 GCTTCCGCTACGACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1206
 DB 651 TGCCACGCGCGACATCTCGGTAATAATGATGATGATGATGATGATGATGATGATGATGATG 710
 QY 1207 CGCGCGCGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1266
 DB 711 CTTACCGTGTCTTACGCGCATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATG 770
 QY 1267 GTTCTGTGACCCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1316
 DB 771 CGCTATGAAACAAGTAAATGCGCGATCTTCTGGAACCGGTTCTGGAAGCATGACAC 830
 QY 1317 -GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1374
 DB 831 CTGCGGCTGCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 890
 QY 1375 TGCATTTGCAAGCTTCCGGAAGATCCGSCACGGAAGG 1413
 DB 891 CCGGTTGCAATATTGCGGATACCCCATGCGGTAGGC 929

RESULT 13

US-10-040-037-1
 Sequence 1, Application US/10040037
 Patent No. US20020102593A1
 GENERAL INFORMATION:
 APPLICANT: Leonard, Jack T.
 APPLICANT: MacDonald, Constance
 APPLICANT: Gabriels, Joseph
 TITLE OF INVENTION: Method for Sequencing Reaction Cleanup
 TITLE OF INVENTION: By Constant Pressure Differential Ultrafiltration
 FILE REFERENCE: MCA-480A
 CURRENT APPLICATION NUMBER: US/10/040,037
 PRIOR FILING DATE: 2001-10-19
 PRIOR APPLICATION NUMBER: 60/154,448
 PRIOR FILING DATE: 1999-09-17
 NUMBER OF SEQ ID NOS: 1
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 1
 LENGTH: 767
 TYPE: DNA

```

; ORGANISM: Escherichia coli bacteriophage lambda
US-10-040-037-1
Query Match 3.9%; Score 64.2; DB 12; Length 767;
Best Local Similarity 95.7%; Pred. No. 8.8e-10;
Matches 66; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1457 GAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGACGCGATGCAAGCTTGGC 1516
|||||
Db 9 GAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGACGCGATGCAAGCTTGGC 68
|||||

QY 1517 TGTATTGGC 1525
|||||
Db 69 TGTATTAGTC 77
|||||

RESULT 14
US-10-013-036-3
; Sequence 3, Application US/10013036
; Patent No. US20020128186A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jeffrey
; TITLE OF INVENTION: Antimicrobial Polypeptide, Nucleic Acid, and Methods of Use
; FILE REFERENCE: 00-1323H
; CURRENT APPLICATION NUMBER: US/10/013,036
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 09/361,900
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: 08/871,924
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Streptococcus mutans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(782)
; OTHER INFORMATION:
; NAME/KEY: -35 signal
; LOCATION: (177)..(182)
; OTHER INFORMATION:
; NAME/KEY: -10 signal
; LOCATION: (191)..(196)
; OTHER INFORMATION:
; NAME/KEY: RBS
; LOCATION: (218)..(224)
; OTHER INFORMATION:
US-10-013-036-3

Query Match 3.8%; Score 63; DB 12; Length 1323;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1454 TAGGAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGACGCGATGCAAGCTT 1513
|||||
Db 912 TAGGAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGACGCGATGCAAGCTT 971
|||||

QY 1514 GGC 1516
|||||
Db 972 GGC 974
|||||

RESULT 15
US-09-179-536B-267/c
; Sequence 267, Application US/09179536B
; Patent No. US20020042112A1
; GENERAL INFORMATION:
; APPLICANT: Hubert K ster
; David M. Lough
; Guobing Xiang
; TITLE OF INVENTION: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY
```

```

;
; NUMBER OF SEQUENCES: 320
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/179,536B
; FILING DATE: 26-Oct-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20444
; FILING DATE: 06-NOV-1997
; APPLICATION NUMBER: 08/947,801
; FILING DATE: 08-Oct-97
; APPLICATION NUMBER: 08/933,792
; FILING DATE: 19-Sep-97
; APPLICATION NUMBER: 08/787,639
; FILING DATE: 23-Jan-97
; APPLICATION NUMBER: 08/786,988
; FILING DATE: 23-Jan-97
; APPLICATION NUMBER: 08/746,055
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/746,036
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/744,590
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/744,481
; FILING DATE: 06-No. US20020042112A1-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24736-2004B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8400
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 267:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 267:
US-09-179-536B-267

Query Match 3.7%; Score 61.6; DB 10; Length 103;
Best Local Similarity 88.2%; Pred. No. 1.7e-09;
Matches 67; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1454 TAGGAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGACGCGATGCAAGCTT 1513
|||||
Db 80 TAGGAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGACGCGATGCAAGCTT 21
|||||

QY 1514 GCGTGTGTTGGCGGAT 1529
|||||
Db 20 GGCACCTGGCGGTCGTT 5
|||||

Search completed: January 16, 2003, 09:42:30
Job time : 96 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2003, 07:25:01 ; Search time 1687 Seconds
(without alignments)
15869.081 Million cell updates/sec

Title: US-09-830-111A-1
Perfect score: 1653
Sequence: 1 ttttgggggtgaaaagtc.....atgccgaactcagaagtga 1653

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estnu.*
- 4: em_estnu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_esti.*
- 9: gb_esti.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	165.4	10.0	561	9	AA901612	AA901612 NCM3A1T3
2	160.4	9.7	656	14	BQ994487	BQ994487 QGF7P20.Y
3	152.8	9.2	843	17	CNS06X9P	AL419459 T3 end of
4	148.4	9.0	690	13	BI722198	BI722198 1031060E0
5	145.2	8.8	986	17	CNS07DTL	AL440911 T7 end of
6	142.6	8.6	621	9	AA538334	AA538334 VJ03F03.F

7	141.2	8.5	597	14	BQ295774	BQ295774 1091042F1
8	137	8.3	949	9	AL582288	AL582288 AL582288
9	136.6	8.3	599	14	BM786570	BM786570 K-EST0065
10	134.2	8.1	897	12	BG252364	BG252364 602365955
11	133.6	8.1	704	13	BQ399049	BQ399049 BQ399049
12	133.2	8.1	735	12	BG125274	BG125274 EST470920
13	129.8	7.9	1112	11	AV108392	AV108392 Zea mays
14	129	7.8	535	12	BF610448	BF610448 NXS1_058
15	125.4	7.6	659	12	BG858032	BG858032 1024055D0
16	124.8	7.5	485	14	BQ142681	BQ142681 123 Metar
17	123.6	7.5	672	10	AW221898	AW221898 EST398709
18	123.2	7.5	650	13	BQ258657	BQ258657 BQ258657
19	120.8	7.3	682	10	AV784831	AV784831 AV784831
20	120.6	7.3	570	14	BQ041590	BQ041590 g2d4all.Y
21	120.6	7.3	665	13	BM409824	BM409824 EST584151
22	120.2	7.3	769	14	BQ853525	BQ853525 QGB20L24.
23	120.2	7.3	771	14	BQ853490	BQ853490 QCB20K12.
24	119	7.2	622	13	BM408807	BM408807 EST583134
25	119	7.2	681	14	BM824056	BM824056 K-EST0095
26	117.8	7.1	519	13	BI423817	BI423817 saf18d11.
27	117.2	7.1	455	13	BI708490	BI708490 tt48d07.Y
28	115.6	7.0	1203	11	AY105886	AY105886 Zea mays
29	114.4	6.9	601	10	BE432562	BE432562 EST399091
30	114	6.9	572	10	BE460050	BE460050 EST415342
31	114	6.9	598	14	BQ819659	BQ819659 1030079E0
32	114	6.9	779	14	BQ018783	BQ018783 UI-H-DH1-
33	113.6	6.9	574	10	BE436140	BE436140 EST407218
34	110.8	6.7	752	14	BQ743828	BQ743828 WHE4108.F
35	108.4	6.6	558	13	BJ488122	BJ488122 BJ488122
36	108.2	6.5	621	9	AI485604	AI485604 EST243925
37	108.2	6.5	629	14	BQ624777	BQ624777 USDA-FP.0
38	107.2	6.5	559	12	BF112700	BF112700 EST440290
39	105.8	6.4	984	13	BI756998	BI756998 603028715
40	105.6	6.4	544	10	AW096482	AW096482 EST289662
41	105.6	6.4	554	12	BG096598	BG096598 EST461117
42	105.6	6.4	658	10	AW093415	AW093415 EST286595
43	105.6	6.4	740	14	BQ858515	BQ858515 QGC10H07.
44	105.2	6.4	676	13	BM139431	BM139431 Hc_ad 23D
45	105	6.4	561	12	BF112719	BF112719 EST440309

ALIGNMENTS

RESULT 1	AA901612	NCM3A1T3	Mycelial Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
LOCUS	AA901612	NCM3A1T3	Mycelial Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
DEFINITION	AA901612	NCM3A1T3	Mycelial Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
ACCESSION	AA901612	NCM3A1T3	Mycelial Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
VERSION	AA901612.1	GI:3045589	Mycelial Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
KEYWORDS	EST.		Mycelial Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
SOURCE	Neurospora crassa		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
ORGANISM	Neurospora crassa		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
REFERENCE	1 (bases 1 to 561)		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
AUTHORS	Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L., Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueves, E., Cushing, T., Erett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R., Ortega, J., Pavlova, I., Perea, J., Todisco, S., Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig, D.O.		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
TITLE	Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
JOURNAL	Fungal Genet. Biol. 21, 348-363 (1997)		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
MEDLINE	97435549		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001
COMMENT	Contact: Natvig, D.O./Nelson, M.A. Department of Biology University of New Mexico Castetter Hall, Albuquerque, NM 87131, USA Tel: 505 277 3411 Fax: 505 277 0304		Neurospora crassa	561 bp	mRNA	linear	EST 10-JAN-2001

Email: ngs@biology.unm.edu

Putative HEXAMERITL PYROPHOSPHATE SYNTHETASE PRECURSOR (HPS) based upon BLASTX search. Deposited in GSDb at the National Center for Genome Resources with accession GSDb:S:1147848

Seq primer: T3.

FEATURES

Source

Location/Qualifiers
1..561
/organism="Neurospora crassa"
/strain="74-OR23-IV A (FGSC 2489)"
/db_xref="taxon:5141"
/clone="NM3A1"
/sex="Mating type A"
/tissue_type="Mycelium"
/dev_stage="E. coli"
/lab_host="E. coli"
/note="Vector: pBluescript SK (-); Site 1: EcoR I; Site 2: Xho I; 2% sucrose for 24 hours. cDNA directionally cloned into pBluescript SK(-) using the Uni-ZAP XR vector system (Stratagene, La Jolla, CA)."
BASE COUNT 109 a 204 c 127 g 121 t
ORIGIN

Query Match 10.0%; Score 165.4; DB 9; Length 561;
Best Local Similarity 63.4%; Pred. No. 4.9e-33;
Matches 301; Conservative 0; Mismatches 171; Indels 3; Gaps 3;

OY ATATCTCGCCTCGCAAGG-CGGTGGCTGAGATCAGAGATGATCCATGACATCA 715
DB ACATCTCTCTCCAGCGAGAGACTGCCGAGATCACCGACTGATCATCCGCTCG 146
OY TCTCTCCAGCAGCAGTATGACGCTTCGAGACCAAGCAAGCACCACATCCGAAAC 775
DB CTCTTCAAGAGATGATGATCATCTCGAGTCCGCGCGCGCTCTCCGCAAC 206
OY 776 C-AGCATTCGGAACCAAGATGCGATTTGGCTGTGATTTCTTTGGAGCGGCGTC 834
DB CTGCAATTCGGAACCAAGATGCGCTCTGCTGTGATCTTCTTTGGAGAGCGCTC 266
OY 835 TGTTSCATTCGAGAGATTCGCAATCCGAGGTGATGAGCTTTGGCTACTGTTATTC 894
DB CGTGCCTCTGCGCCCTCGCCGCAAGTCTGAGCTTTCTTCTGACCTGATTC 326
OY 895 AAATCTGTTGAGGAGAGATTCAGTGA-AAAATCTGTTGATGATGCGATTGAG 953
DB CAACCTTGCAGGAGGAGATTCAGTCAAGAAACAGCGCTCGCGCAAAAAAAGC 386
OY 954 CTAGCGCAGCAGGAAAGCTTCATTAATTTCAGAAAGCTTAATTGAAGACTGCT 1013
DB CCCAATGATCGAGAGAGACTCTCACTAATCCAAAGACTAATCAAGACCGCTT 446
OY 1014 CTTGATTCGCAAGTCTGTCAGAGCAAGTCTGAGGTGATGAGCTGAGGTTG 1073
DB CCGTATCAGCAAGACTCTGCGCTCTGCGCTGAGGGGCGCTATGCGGAGCG 506
OY 1074 CTGATGCTGCTTATCTTACGAGAAAGCTTTGGCTATTCAGATCGTGA 1128
DB TGATGCTGCTTATCTTATGCAAGAAAGCTTGGCTTCCAGGTGTCGA 561

RESULT 2

BO994487 656 bp mRNA linear EST 21-AUG-2002
LOCUS BO994487
DEFINITION OGF7F20.Y9.abi OG FRGHU lettuce serriola Lactuca sativa cDNA clone
OGF7F20 mRNA sequence.

ACCESSION BO994487
VERSION BO994487.1 GI:22414022
KEYWORDS EST.

SOURCE

Lactuca sativa.
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; Asteraceae; Lactuceae;

REFERENCE

1 (bases 1 to 656)

AUTHORS

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., Van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Composite Genome Project
http://compgenome.ucdavis.edu/
Unpublished (2002)

JOURNAL

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
singletons, see http://cgdb.ucdavis.edu/ for details.
Plate: OGF7 row: F column: 20.

FEATURES

Source

Location/Qualifiers
1..656
/organism="Lactuca sativa"
/cultivar="L. serriola"
/db_xref="taxon:4236"
/clone="OGF7F20"
/lab_host="E. coli"
/note="Vector: pBRCDNA5flab; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformationally made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG-TISSUE-flowers pre-fertilized
TAG-SEQ=GGTTGACGGG"

BASE COUNT 195 a 125 c 152 g 182 t 2 others
ORIGIN

Query Match 9.7%; Score 160.4; DB 14; Length 656;
Best Local Similarity 54.0%; Pred. No. 1.2e-31;
Matches 363; Conservative 0; Mismatches 288; Indels 21; Gaps 1;

OY 728 GACGTTATCGACGCTTCGAGACAGCAAGAACACCATCCGGAACCGACATTCGGA 787
DB 1 GATGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
OY 788 AACAGATGCGGATTTGGCTGTGATTTCTTTGGAGCGGCGTCTGTGATTCGCG 847
DB 61 AATAAGCTTCAGATATGCTGAGAGATTTCTTCTAGAGCTGATTAACCTCGCC 120
OY 848 AGTTCGCCCATCCGAGAGTATGAGCTTTGGCTACTGTTATTCGAACTTGGTTAG 907
DB 121 TCTTTAAAAACACAGAGCTGTATCACTAATTAACAGCTGTGAGCACTTGTGACA 180
OY 908 GAGAGCTTCATGAGTGAATAATCTGTTGATGATGATGATGATGATGATGATGATG 967
DB 181 GGTGAAACATGCAATATAGACGTCAGACAGACGCTAT----- 222
OY 968 GAAACGTTGATTAATTTGACAGAGACTTAAGAGTGGCTGCTGTTGATTCGCAAG 1027
DB 223 ---AGCATGATTAATTTGACAGAGAGTATTAAGAGTGCATCTTTAATTTCAAC 279
OY 1028 TCGTGCAGAGCAAGTCCCTTCTGAGTGTCTACGCTGAGGTTGCTGATGCTGTTAT 1087
DB 280 AGCTGCAATTCATTCACCTTTACCGGCCAACTATCTGAGTGGCTGATGCTGCAAT 339
OY 1088 GCTTACGAGAGAACCTTTGTTGGCATTCAGATCGCGACGACATGCTGACATACAC 1147
DB 340 GAATACGAGAAAAATCTTTGTTGGCATTCGAATTAATGATGATGATGATGATGAT 399


```
QY 1148 GTCTCGCTACCGACTCGGTAAAGCCGCGGTGCAGACCTCCAGTCTCGTCTCGCCACC 1207
Db 400 GGCACATCATCTCTAGGGAAGGGTGCATTATCTACATTACCATTGGAATGTGACA 459
QY 1208 GCCTCGGCGCTCTTCGCATGGAAGCACACCGCGAGTCCGCTCCCATGATCAAGCGCAAG 1267
Db 460 GCTCCAAATATATGCAATGGAGGAGTTTCCCGAACTACGTACAGTTGTCGACAGGGC 519
QY 1268 TTCTCTGACCCAGAGACGTGCGAGCGGTGCGACGGAGTTGTCGAGAAAAGTGTGAAATG 1327
Db 520 TTGGATGACCCCTGAAAAACGTGAATCTTGCTCTGGAGTATCTTGAAAAAGTCNTGGGATA 579
QY 1328 GAGAAGCAGAGACCTTGGCGGAGGAGTATGCCAGAAAGCGTGGATGCAATTCGACG 1387
Db 580 CNAAGAAACAGAACTGGCAGCAAGCATGCCAGCGCTGCGTCTGCTGCAATTTGATTCA 639
QY 1388 TTCCCGGAGAGT 1399
Db 640 TTGCTGAAAT 651

RESULT 3
CNS06X9P/c
LOCUS
DEFINITION
T3 end of clone AX0AA039F06 of library AX0AA from strain CBS 7064
of Pichia farinosa, genomic survey sequence.
ACCESSION
AL419459
VERSION
AL419459.1 GI:12202637
KEYWORDS
GSS.
SOURCE
Pichia farinosa.
ORGANISM
Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
1 (bases 1 to 843)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 843)
de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Potier,S.
Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
sorbitophila
FEBS Lett. 487 (1), 87-90 (2000)
20584725
11152890
3 (bases 1 to 843)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
Keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..843
/organism="Pichia farinosa"
/strain="CBS 7064"

FEATURES
Source
```

```
misc_feature
complement(<309..>839)
/clone="AX0AA039F06"
/clone_lib="AX0AA"
/feature="end : T3"
/feature="similar to Saccharomyces cerevisiae ORF YBR003w [
COO1 : hexaprenyl pyrophosphate synthetase precursor ]"
/feature="not experimental"
BASE COUNT 350 a 143 c 139 g 192 t 19 others
ORIGIN
Query Match 9.2%; Score 152.8; DB 17; Length 843;
Best Local Similarity 57.8%; Pred. No. 1.5e-29;
Matches 288; Conservative 2; Mismatches 194; Indels 0; Gaps 0;
QY 965 CAGGAAACGTTTCGATTACTATTTGCAGAAAGACTTACTTGAAGACTCGTCTTGATTC 1024
Db 728 CAAAGCAGCATTTGAATACTATTTGCACAAAGACTTACTTGAAGAACTCGTCTTTAATGCT 669
QY 1025 AAGTCGTGACAGCAAGTGGCTTCTGGGTGGTGTCTACGCTGAGGTGCTGATGCTGCT 1084
Db 668 AACTGTCCAGGCGAGTGCAGTGTTCAGTGGAGCGCAAGATGATGTTATTGGAATTGC 609
QY 1085 TATGCTTACGGAAGAACTTGGTTTGGCATTCAGATCGTCGACGACATGCTCGACTAC 1144
Db 608 TATGAGTTTGGCGGTAAATTTGGGTTTATGCTTCCAAATAGTAGACACATATTAGACTAC 549
QY 1145 ACCGTCTCCGCTACCGACCTCGGTAAAGCCGCGGTGCAGACCTCCAGCTCGGTCTCGCC 1204
Db 548 ACCTCAAGCGCAATACTTTTGGTAAACCTAGTCAAGCAGATCTTAAGTTAGTTAGCA 489
QY 1205 ACCGCGCGGCGCTCTTCGCATGGAAGACACCGCAGCTCGGTCCCATGATCAAGCGC 1264
Db 488 ACTGCTCCAAATTTTATTCGCTTGAAGAGAGAACCAACTTGGCCAAATTAATCGCCAGG 429
QY 1265 AGTTCTCTGACCCAGAGACGTCGAGCGTGCAGCGAGTTGTCGAGAAAGTATGGA 1324
Db 428 AAATTTAGCGAGGAAGAGACGTTGAACCTGCGCACTAATGCCGTGTAGAAATTTGATGGT 369
QY 1325 TTGGAAGACGAGAGCGCTTGGCGGAGGAGTATGCCAGAAAGCGTTGGATGCAATTCGG 1384
Db 368 TTAGAGAGAGAGAGAAATGGCTGAGGAGTACCGTTTCAAGGCTTTAGAAAACCTGAGA 309
QY 1385 ACCTTCCCGAGAGTCCGCGACGGAAGGCTTTGAGAGCGTTGAC 1428
Db 308 TTCTTGTGATTCCGWTGCCWTKTATTTCTTTGTTTTTTTAC 265

RESULT 4
BI722198 690 bp mRNA linear EST 19-SEP-2001
LOCUS 1031060E08.y1 C. reinhardtii CC-1690, Stress II (normalized),
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI722198
VERSION BI722198.1 GI:15697893
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 690)
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
```

FEATURES Email: chausereduke.edu.
source Location/Qualifiers

1. 690
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_11b="C. reinhardtii CC-1690, Stress II (normalized

), Lambda Zap II"
/note="Vector: Bluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the Lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 113 a 225 c 248 g 104 t
ORIGIN

Query Match 9.0%; Score 148.4; DB 13; Length 690;
Best Local Similarity 55.5%; Pred. No. 2e-28;

Matches 321; Conservative 0; Mismatches 226; Indels 21; Gaps 1;

641 GAGGAGACAGCTGAAATCTTCGCTCGCAACGCGGTTGGCTGAGATCAGAGATG 700
134 GCGGAGACCCGATGATCGCGGCGGAGCAGCGCTCGCGAGATCGCGAGCTC 193
701 ATTCATGACAGATCATCTCTCCAGACAGCTTATCGACGCTTCCGAGCCAGACGAAAC 760
194 ATCCACGTGGCAGCTCTTACGATGATGATGATGATGATGATGATGATGATGATGATG 253
761 GCACCATTCGGAACCAAGGATTCGGAACCAAGATGCGATTTGGCTGATTTCTTG 820
254 GTGCTGTGCTCAATACAGGCTGGGCAACAGCGCCATTCGCGGGGAGATTCTTG 313
821 TTGGAGCGGGCTGTGCTGATTCGAGAGTTGGCAATCCGAGGATGATGATGATGATG 880
314 CTGGCGGGGCTCGTGAACGCTGCGCTGCGCAACAGTGAATCGTGAATGATGATG 373
881 GCTACTGTTATGCAAACTGTTGAGGAGATTCATGCACTGTTGAAAATATCTGTTGAT 940
374 AGTCAGGTGCTGAGACACTGCTGCGGCGAGATCATGAGATGACCGCCACAGGAG 433
941 GATCGATTGAGGCTAGCGGCGAGCAGGAAAGTTTCATTTCTATTGCAAGACTTAC 1000
434 CAGTCTCTGAC-----TGAAGCACTACCTGGCCAAACCTAC 472
1001 TTGAAGCTGCGCTCTGATTGCCAAGTCGTGAGAGCAATGCGCTTGGGTGCT 1060
473 TGAAGAGAGCAGCCTCATGGAACAAGCTCGGCTCGGCGGCTGCGGAGTGC 532
1061 AGGCTGAGGTGCTGATGCTGCTTATGCTTACGGAAGAACTTGTGTTGGATTTCAG 1120
533 GCGCGAGAGGTGCTGAGATGCGGTGAGCTACGCGCCCACTGGGAGATCGCTTCAG 592
1121 ATGCTGAGCATGCTGCTGATACCGCTACCGGCTCGGTAAGCCGCGCGT 1180
593 GTGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652
1181 GCAGACCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1218
653 AACGACATGCGCAGTGGGCTGGCCACGCGCGGTGCT 690

RESULT 5

CNS07DTL/C
LOCUS CNS07DTL 986 bp DNA linear GSS 08-JUL-2001
DEFINITION T7 end of clone BD0A015F02 of library BD0A from strain CBS 94 of
Candida tropicalis, genomic survey sequence.
ACCESSION AL440911
VERSION AL440911.1 GI:12224322
KEYWORDS GSS.
SOURCE Candida tropicalis.
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE

1 (bases 1 to 986)
Soucier, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bollon-Fukuhara, M., Bon, E., Broctier, P., Casaregola, S.,
de-Montigny, D., Dujon, B., Durand, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neugebauer, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, M., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL

MEDLINE
PUBMED
20584711
1152876

REFERENCE

2 (bases 1 to 986)
Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and
Dujon, B.
Genomic exploration of the hemiascomycetous yeasts: 16. Candida
tropicalis
FEBS Lett. 487 (1), 91-94 (2000)

JOURNAL

MEDLINE
PUBMED
20584726
1152891

REFERENCE

3 (bases 1 to 986)
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web :
http://www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia soboloffii,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

COMMENT

FEATURES
source
1. 986
/organism="Candida tropicalis"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone="BD0A015F02"
/clone_11b="BD0AA"
/note="End : 77"
complement(<6. >899)
/note="similar to Saccharomyces cerevisiae ORF YBR003w [
COO1 : hexaprenyl pyrophosphate synthetase precursor]"
evidence=not experimental

BASE COUNT 298 a 209 c 141 g 337 t 1 others
ORIGIN

Query Match 8.8%; Score 145.2; DB 17; Length 986;
Best Local Similarity 52.2%; Pred. No. 1.9e-27;

Matches 458; Conservative 0; Mismatches 318; Indels 102; Gaps 2;

671 CAAGCGGCTGGCTGATGATCAAGAGATGATGATGATGATGATGATGATGATGATGATG 730
887 CAAAGAGATTTGGCTGAGATTGTGGAATGATGATGATGATGATGATGATGATGATGATG 828
731 GTTATGAGCTTTCGAGACGAGAGAAAGCAATCGGAAACAGGCAATTCGGAAC 790
827 GTATGATTTGCTGATTCAGAAAGGTAGACCAAGCGGAAATTTGCAAAAT 768

Qy	791	AAGATGGCGA	TTTTGGCTGGTGA	TTTTCTTTGGGACGGGGCTCTGTTCGATTTGGCGAGG	850
Db	767	AAAATGGCTGTTT	TAGCAGGTGA	TTCTTATTGGGAAGAGCTTCTGTGTCCATCGCCAGA	708
Qy	851	TTGGCRAATCCGGAGGTG	CATTGAGCTTTTGGCTACTGTTATGCAAACTTGGTTGAGGA	910	
Db	707	TTAAGAAACCCAGAGATT	TATTGAATGTGTGAGTACTACCATTTGCCAATTTGGTTGAAGGG	648	
Qy	911	GAGTTTCATGCACTT	GAAAAAATACTGTT	-----	937
Db	647	GAATTAATGCAAT	TGAAGAAATACCGTTATACAAACAAATGATGATCAAAATAGTAAT	588	
Qy	938	-----	-----	-----GATGAT	943
Db	587	GATGGTGAAGTT	AAAGAAATCCCTCGTCTACAGGTAAAGTTCCAACCAAACTCCATGAT	528	
Qy	944	CGGATTGAGGCTA	CGGCGACGGAGAAACG	-----TTCGATTACTATTTG	988
Db	527	TATTCTGTTAGAACT	TGTGACCATGAAACAAATGTGTGCTGCAATTTGAAATATTATTG	468	
Qy	989	CAGAAGACCTT	ACTTGAAGACTCGCTTGATGTCGAAGTCGTGCAGAGCAAGTCGGCTT	1048	
Db	467	CACAAAACATATT	TGAAACAGCCCTCAATTAATGTCCAAGTCTTGTCGTCAGCAGCTGTT	408	
Qy	1049	CTGGTGTGTGCT	CAGCCTCAGGTTGCTGATGCTGCTTATGCTTACGGAAGGAACCTTGGT	1108	
Db	407	TTAAGTGTGTCT	CAAGATGATGTCATTGAAACCTGTTATAGTTTGGGAAGAAATTTGGGA	348	
Qy	1109	TGGCAATTCAGATG	TCGACACATGCTCGACTACACCGTCTCCGCTACCGACCTCGGT	1168	
Db	347	TTATGTTTCCAAAT	AGTAGTCGATGATATTGGATTATCTAGTAGTGATAAAGCTATTGCT	288	
Qy	1169	AAGCCCGCGGTGC	AGACCTCAGCTCGGTCTGCGCACCGCGCGGCCCTCTTTCGATGG	1228	
Db	287	AAACCAAGTCAAG	CTGATTGGAATTTGGATTGGCAACTGCTCCAAATATTGTATGCATGG	228	
Qy	1229	AAGCACACGCGAG	CTCGGTCCCATGATCAAGCGCAAGTCTCTGACCCAGGAGACGTC	1288	
Db	227	AAAGAAGAAC	CAGCATTAGGTGAATTAATTGGCAGAAAGTTCAGTCAACCAAGGTGATGTT	168	
Qy	1289	GAGCGTCACGCG	AGTTGGTTCGAGAAAAAGTATGGATTGGGAAGACGAGAGCCCTTGGCG	1348	
Db	167	GAATAGCCAGAG	CTGTTGATAAATATAATGSGGTTGAAAAACAAGAGAAATGGCC	108	
Qy	1349	GAGGAGTATGCC	AGAAGCGGTGGATGCAATTCGACGTTTCCGGAGAGTCCGCGACGG	1408	
Db	107	ACCATGTTTGT	CATGAAGCATTTGAAGAAATTAAGATGTTTACCTGACTCTCAAGCTAGA	48	
Qy	1409	AAGCTTTGGAG	CAGTTGACGCAAGGTGTGACTAG	1446	
Db	47	AGTGCATTGGA	ATATTGACCAATTTCTGTTTAAACGAG	10	

RESULT 6	AA538334	621 bp	mRNA	linear	EST 29-JUL-1997
LOCUS	AA538334				
DEFINITION	vJ03f03.r1 Barstead mouse pooled organs MPRB4 Mus musculus cDNA clone IMAGE:920869 5', similar to SW:COO1 YEAST P18900 HEXAPRENYL PYROPHOSPHATE SYNTHETASE PRECURSOR ; mRNA sequence.				

VERSION	AA538334.1	GI:2284327
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 621)	
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	

```

Db      475 GAGATGTGGACAGACGACAGATATGATTACAGAGTATGGCTGACGAAACAACCT 534
Qy      1341 CTTGGCGGAGAGTATGCCAGAAAGCGTTGATGCAATTCGACCTCCCGAGAGTC 1400
Db      535 ACCTCGCCACGACATCTGCCCAAGAGCTGTGAGAGATCAGAACTTAGACATCTA 594
Qy      1401 CGGCACGAAAGCTTTGGAGCAGTT 1425
Db      595 CAGAAAGGACGCCCTCATTCAGCT 619

RESULT 7
LOCUS   BQ295774          597 bp      mRNA      linear      EST 16-MAY-2002
DEFINITION 1091042F12.v1 1091 - Immature ear with common ESTs screened by
            Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION BQ295774
VERSION   BQ295774.1  GI:20811296
KEYWORDS  EST.
SOURCE    Zea mays.
           Zea mays.
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
           clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 597)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
          University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
          Department of Biological Sciences
          Stanford University
          855 California Ave, Palo Alto, CA 94304, USA
          Tel: 650 723 2227
          Fax: 650 725 8221
          Email: walbot@stanford.edu
          Plate: 1091042 row F column: 12.
          Location/Qualifiers
            source
              1..597
                /organism="Zea mays"
                /cultivar="OH43"
                /db_xref="taxon:4577"
                /clone_lib="1091 - Immature ear with common ESTs screened
                by Schmidt lab"
                /cissue_type="Inflorescence meristem - floral organ
                primordia"
                /dev_stage="0.5 cm to 2 cm"
                /lab_host="Stratagene XLDR"
                /note="Organ: Immature ear; Vector: PAD-GAL4; Site_1:
                EcoRI; Site_2: XhoI; RNA from library 606 was filtered for
                common ESTs found in 606."

BASE COUNT      155 a      125 c      148 g      169 t

Query Match      8.5%; Score 141.2; DB 14; Length 597;
Best Local Similarity 55.9%; Pred. No. 1.5e-26;
Matches 322; Conservative 0; Mismatches 233; Indels 21; Gaps 2;

Qy      670 GCAACGGCGGTTGGTGTGATCAGGAGATATCCATGACATCATCTCCCGACGAGA 729
Db      3 GCAACAGAACATTTGCTGATTAAGTGAATGATTCATGTCGCAAGCCTTCTGATGATGA 62
Qy      730 CGTATGACACCTTCCGAGACGAAAGCAACATCCGAAACAGAGCATTCGAAA 789
Db      63 TGTCTTGATATGCTGATTAAGCGGTGTGATGATTCATGATCTCATCATGGGAA 122
Qy      790 CAAGATGGCAATTTGGCTGTGATTTCTTTGGAGAGGGCGTCTGTGATTCGGCAG 849
Db      123 CAAGCTTTCTGTGCTGGCTGTGATTTCTTCTGTGAGACATGTGGCCCTTGACAGC 182
Qy      850 GTTGGCAATCCGAGGATGATGAGCTTTGGCTAAGCTTTTGGCAACTTGGTTGAGGG 909

```

```

Db      183 ACTGGGAACACAGAGGTGTTCTCTAATGSCAATGCTGTAGAACATCTAGTACTGG 242
Qy      910 AGAGTTCAATGCACTGTAATAATTAATCTGTGATGATGCGATTGAGCTACGGCAGCAGA 969
Db      243 TGAACCTATGACAGATCTCAACAAGCAGAGA-----GCAACGGGAGAGCAGT- 288
Qy      970 AACGTTGATTTCTAATTTGCAAGAAGCTTACTGAAGACTGCTGCTGATTTGCCAAGTC 1029
Db      289 -----GAGTACTACTGTCAGAAAGACATTAACAAACGGCATATTGATATCAAAATG 341
Qy      1030 GTGACAGCAAGTGGCGCTTCTGGTGTGTGTCACGCTGAGGTGTCGATGCTTATGTC 1089
Db      342 TTGCAAGGCTGTGCTATTTCTTGACAGGACACAACTAGGCTGTGGCTTGCAATAGA 401
Qy      1090 TTACGGAAGAACTTTGTTGGCATTCGATCGTCGACGACATGCTGACACTACCGT 1149
Db      402 ATATGCTCAAAACCTGGGCTGACCTCCAGTTAATGATGATGTTGATTTGACCGG 461
Qy      1150 CTCGGCTACCAAGCTCGGTAGACCCGCGGTGACAGACCTCCAGTCTGCTGCGACCGC 1209
Db      462 CACCTTGATCCCTTGGGAGGGTTCAATGTTGATATTCGCCACGAAATTATTACTGC 521
Qy      1210 GCCGGCCTCTTGCATGAGAACACACCGCCGAGCT 1245
Db      522 CCGATGCTAATATGCGATGAGGAATTCACCAACT 557

RESULT 8
LOCUS   AL582288/c          949 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION AL582288 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0D005YM10 3
            prime, mRNA sequence.
ACCESSION AL582288
VERSION   AL582288.1  GI:12950123
KEYWORDS  EST.
SOURCE    human.
           human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 949)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 Evry cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
          Location/Qualifiers
            source
              1..949
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0D005YM10"
                /clone_lib="LTI_NFL010_BC2"
                /sex="male"
                /cissue_type="B cells from Burkitt lymphoma"
                /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                was primed with a NotI-Oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : filiang@lifetech.com URL :
                http://fulllength.invitrogen.com"

BASE COUNT      285 a      192 c      185 g      263 t      4 others

Query Match      8.3%; Score 137; DB 9; Length 949;
Best Local Similarity 51.5%; Pred. No. 2.8e-25;
Matches 354; Conservative 1; Mismatches 311; Indels 21; Gaps 1;

Qy      781 ATTGGAAACAGAGTGGCAATTTGGCTGTGATTTCTTTGGAGACGGGCGTCTGTTGC 840

```

```
Db 945 ATCTGGTGAAGAGCTGTTCTTCTGGAGATTAAATTTCTTCTCAGCATCTATAGC 886
Qy 841 ATTGGCAGGTGGCGAATCCGAGGTGATGAGCTTTTGGCTACTGTTATTGCAAACTT 900
Db 885 TCTGGCAGCAATTTGGAATACAACTGTTATCTATTATTAACCAAGTTATTGAAGATTT 826
Qy 901 GGTGTAGGAGAGTTCATGTCAGTGTGAATAAATCTGTTGATGTCGATTCGAGCTACGGC 960
Db 825 GGTGCGTGGTGAATTTCTTCAGCTCGGGTCAAAAGAAATGA-----784
Qy 961 GACGCAAGAAACGTTGCTATTACTATTTCGCAAGAGCTTACTTGAAGAGCTCGTCCCTTGAT 1020
Db 783 ---GAATGAAGATTGTCACACTACCTTGTGAGAGACATTCAAGAGAGCCCGACCTGAT 727
Qy 1021 TGCAGTCTGTGAGAGCAAGTGGCTTCTGGTGGTGGTGTACGCTGAGTGTGCTGATGC 1080
Db 726 AGCCAACAGTGTGAAGCAGTCTCTGTTCTAGGATGTCCCGACCCAGTGGTGCATGAGAT 667
Qy 1081 TGCTTATGCTTACGGAGGAACTTGGTTTGGCATTCAGATGTCGACGACATGCTCGA 1140
Db 666 CGCTATCAGTACGGAAGAAATGTAGGAATAGCTTTTCAGCTAATAGATGATGATTGGA 607
Qy 1141 CTACACGCTCTCCGCTACCGACCTCGGTAAGCCGCGGTGCAGACCTCCAGCTCGGTCT 1200
Db 606 CTTCACTCTGTGTTGACAGATGGCAACCAACATCAGTGTGATCTGAAGCTCGGTT 547
Qy 1201 CGCACGCGCGCGGCTCTTTCGATGGAAGCAGCCGCGAGCTCGGTCCCATGATCAA 1260
Db 546 AGCCACTGGTCTGCTGCTGTTGCTGTCAGCAGTTCACAGAAATGAATCTATGATCAT 487
Qy 1261 GCGCAGTCTCTGACCCAGAGAGCTGAGCGTGCAGCGGATTTGGTCGAGAAAGTGA 1320
Db 486 GCACCGGTTTCAGTTCGCTGAGATGTAGACAGAGCTCGACAGTGTACTACAGAGTGA 427
Qy 1321 TGGATTGGAGAAGACGAGAGCCCTTGGGGAGGAGTATGCCAGAGAGCGTGTGGATGCAAT 1380
Db 426 TGGTGTGCAACAACACCTTCCTGCGCCAGAGAGCTGCGCATGAAGCAATAGAGAGAT 367
Qy 1381 TCGGAGCTTCCCGAGAGTCCGGCAGGAAGGCTTTGGAGCAGTGTACCGGACAAAGTGT 1440
Db 366 CAGTAAACTTCGACCATCCCGAGAAAGAGATGCCCTCATTCAGCTTTCAGAAATGTACT 307
Qy 1441 GACTAGTCAAGATAGGAATTCAGCT 1467
Db 306 CACAAGAGATAAATGACAACTCTTTCT 280

RESULT 9
BM786570 599 bp mRNA linear EST 05-MAR-2002
LOCUS K-EST0065318 S9SNU601 Homo sapiens cDNA clone S9SNU601-39-C05 5',
DEFINITION mRNA sequence.
ACCESSION BM786570
VERSION BM786570.1 GI:19134802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 599)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
```

```
Plate: 39 row: C column: 05
High quality sequence stop: 599.
Location/Qualifiers
1. 599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S9SNU601-39-C05"
/clone_lib="S9SNU601"
/sex="M"
/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then dephosphorylated
with tabacco acid pyrophosphatase (TAP). The dephosphorylated
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII- digested pME18-FL3 vector. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
BASE COUNT 176 a 124 c 142 g 157 t
ORIGIN
Query Match 8.3%; Score 136.6; DB 14; Length 599;
Best Local Similarity 53.2%; Pred. No. 2.6e-25;
Matches 329; Conservative 0; Mismatches 269; Indels 21; Gaps 1;
Qy 738 ACCTTCCGAGACACGACGACCATCCGAAACAGGCATTCGGAACAGATGG 797
Db 1 ACATGCAAGTTCTCGAAGAGGAAACACACAGTTAATAGATCTGGGTGAAAGAGG 60
Qy 798 CGATTTTGGCTGCTGATTTCTTCTGGGACGGCGTCTGTTGCATTGGCGAGGTGCGCA 857
Db 61 CTGTTCTGCTGGAGATTAAATCTTTCGACATCTATAGCTTGGCAGATTCGAA 120
Qy 858 ATCCGAGGTGATGAGCTTTTGGCTACTGTTATTCGAACTTGGTTGAGGAGATTC 917
Db 121 ATACAACCTGTTATCTATCTATTTTAAACCAAGTTATTGAAGATTTGGTGGTGAATTC 180
Qy 918 TGCAGTTGAAATACTGTTGATGTCGATTGAGCTACGGCGACGACGAAACGTTCC 977
Db 181 TTCAGCTCGGTCAAAAGAAATGA-----GAATGAAGATTTG 219
Qy 978 ATTACTATTTCGAGAAGACTTACTTGAAGACTGCGTCTTGTGTTGCAAGTCGTGCAGAG 1037
Db 220 CACACTACCTTGAAGAGACATTCAAGAAACCCGACGCTGATAGCCACAGTTGTAAG 279
Qy 1038 CAAGTCGCTTCTGGTGGTGTGCTACGCCGTGAGTGTGCTGATGCTGTTATGCTTACGGAA 1097
Db 280 CAGTCTCTGTTCTAGGATGTCCCGACCCAGTGGTGCATGAGATCGCTATCAGTACGGAA 339
Qy 1098 GGAACCTTGGTTGGCAATTCAGATCGTCGACGACATGCTCGACTACACGCTCTCGCTA 1157
Db 340 AAAATGTAGGAATAGCTTTTCAGCTAATAGATGATGATTTTGGACTTTCACCTCGTGTCTG 399
Qy 1158 CGACCTCGTGAAGCCCGCGGTGCGAGACCTCCAGCTCGGTCTCGCCACCGCGCGGCC 1217
Db 400 ACCAGATGGGCAACCAACATCAGCTGATCTGAAGCTCGGTGATGCTGCTGTTCC 459
Qy 1218 TCTTCGATGGAGACACCGCGAGCTCGGTCCCATGATCAAGCGCAAGTCTCTGACC 1277
Db 460 TGTTCCTGTCAGCAGTTCCTCCAGAAATGAATGCTATGATCATGCGACGCTTCAGTTGC 519
Qy 1278 CAGGAGACGTGAGCGCTGCAACGCGAGTTGGTTCGAGAAAGTGTGATTTGGAGAACGCA 1337
```



```

Db 633 GGTGATTATTATTTGGCAAGCAACAGTAGTTTATCAACAATTAGAAATCCAGATGTA 574
QY 869 ATTGAGCTTTTGGCTACTGTTATTGCAAACTGTTGGTGGAGGAGATTTCATCGAGTTGAAA 928
Db 573 ACTGAATGATGTCACACTGCTTGGCCGAGTTGTTGGAGGTGAATTTATGCAAGCAAAA 514
QY 929 AATACGTTTCATGATGCGATTGAGGCTACGGCAGCGAGAAACGTTTCGATTACTATTG 988
Db 513 TCAAAATGGTCTGCTCTTTT-----GATAATTATTTA 481
QY 989 CAGAAGACTTACTTGAAGACTCGCTCTGATTGCCAAGTCGTGCAGACAAGTGCCTT 1048
Db 480 CAAAAACATTATTAAAACTGGTCTTTTAAATTAACAACAGTTGTAGATCTGCTGCTATC 421
QY 1049 CTGGTGGTCTACGCTGAGGTTGCTGCTGCTTATCTTACGGAAGAAACCTTGGT 1108
Db 420 CTTTCTGGTGCAGATTCAAAACATTATAAATATTCAACTGAATTTTGGAAAGAAATTTAGGT 361
QY 1109 TTGGCATTCCAGATCGTCGAGCATCTCGACTACACGCTCTCCGCTACGACCTCGGT 1168
Db 360 TTAGCTTTTCAAAATGTTGATGATTATTATAGATTATACAGGATCAGCTGAAGAATGTGGA 301
QY 1169 AAGCCCGCGGTGCAGACCTCCAGCTCGGTCTGCCACCGCGCCGCTCTTCGCAATG 1228
Db 300 AAGCTACATCAGTTGATTTAACTTTGGGATTAGCAACACACAGTTTATATGCAACC 241
QY 1229 AAGCACACCCGAGTCTGGTCCCATGATCAAGCGCAAGTTCTCTGACCAGGAGAGCTC 1288
Db 240 CAAGAATTCCACAAATPAGAGAAATTAATTAAGAAAGATTTTCAGAGATTGGTGATTT 181
QY 1289 GAGGTGCACGCGAGTTGTCGAGAAAGTGTGATGATGAGTGGAGAGACGAGCGCTTGGCG 1348
Db 180 GAAGAAGCAAGAGGTTGGTGTCTTTAAGTAAAGGTATCGAGAAAACTAGAAAATTTAGCT 121
QY 1349 GAGGAGTATCCAGAGGCGTTGGATGCAATTCGAGCTTCCGGAGAGTCCGGCAGCG 1408
Db 120 ATGAATATTGTAATCGTCAATTCATCTTATTAATAATACCAATCTGAATCAAGA 61
QY 1409 AAGGCTTTGGAGCAGTTGACGCAAGGTTGCTAGTGGTCAAGATAGGAATTCGA 1464
Db 60 GATCTTTTATCATCATTTATCTCATATTGTTTACAAGACAAAGTAAATAAGCA 5

RESULT 12
BG125274
LOCUS EST470920 tomato shoot/meristem Lycopersicon esculentum cDNA clone
DEFINITION CTOF8K7 5' sequence, mRNA sequence.
ACCESSION BG125274
VERSION BG125274.1 GI:12625462
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 735)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Roming,C. and Tanksley,S.
TITLE Generation of ESTs from tomato shoot/meristem tissue
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1. .735
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"

FEATURES
source
```

```

/clone="CTOF8K7"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 200 a 149 c 168 g 218 t
ORIGIN
Query Match 8.1%; Score 133.2; DB 12; Length 735;
Best Local Similarity 59.1%; Pred. No. 2.4e-24;
Matches 259; Conservative 0; Mismatches 158; Indels 21; Gaps 1;
QY 682 GGCTGAGATCAGGAGATGATCCATGCAGCATCACCTCTCCAGACGACGCTTATCGACGC 741
Db 278 GGCTGAGATCAGTGAATGATCCATGTTGCTAGCCTACTTCATGATGATGCTACTGGATGA 337
QY 742 TTCGAGACGACGAAACGACCATCCGGAAACGAGGCAATTCGGAACAAAGATGGCGAT 801
Db 338 TGCTGACACAAGACGTTGGGATAGTCTTTTAAACTTTGTGATGGAAATTAAGCTAGCTGT 397
QY 802 TTTGGCTGCTGATTTCTTGTGGGACGGGCTCTGTTCGATTTGGCAGGTTGGCAATCC 861
Db 398 ACTAGCCGGAGACTTTTGTCTTCCGAGCATGTGTGGCACTTGCCTCTCTTGAAGAACAC 457
QY 862 GGAGGTGATTGAGCTTTTGGCTACTGTTATTGCAAACTTGGTTGAGGAGAGTTTCATGCA 921
Db 458 AGAGGTGTTATGTTCTCTGGCAACTGTTGTGGAACATCTTGTACTGGAGAGACAATGCA 517
QY 922 GTTGAATAAATACCTGTTGATGATCGGATTCGAGGCTACGGCAGCAGGAAACGTTTCGATTA 981
Db 518 AATGACGACTTCTTCTGATGA-----ACGTTGTAGCATGGAGTA 556
QY 982 CTAATTTGCAAGACACTTACTTTGAAGACTGGCTCTTGTATGTCACAGTCGTGTCAGAGCAAG 1041
Db 557 TTATATGCAAGAAACATATTTACAAGACTGTCATCATTTCAATTTCAATAGGTGCAAGCAAT 616
QY 1042 TGGCTTCTGGGTGGTGTACCGCTCAGGTTGCTGATGCTGCTTATGCTTACCGAAGGAA 1101
Db 617 TGCACTACTTGTGGGCACTAGTCTCGAAGTCTCCGCTGGCTTTTGTGACTACGGGAAAA 676
QY 1102 CTTGCTTTGGCATTCCA 1119
Db 677 TCTGGGATTGGCATTTCA 694
RESULT 13
AY108392
LOCUS Zea mays PCO073322 mRNA sequence.
DEFINITION Zea mays PCO073322 mRNA sequence.
ACCESSION AY108392
VERSION AY108392.1 GI:21211470
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1112)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
JOURNAL Overgo Probes
REFERENCE Unpublished (2002)
AUTHORS Coe,E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
Location/Qualifiers
```

source 1. 1112
 /organism="Zea mays"
 /db_xref="MaizeDB:634028"
 /db_xref="taxon:4577"
 /clone="PC007332"
 /clone_1ib="Maize Mapping Project/DuPont Cornsensus library"
 /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
 BASE COUNT 344 a 221 c 248 g 299 t
 ORIGIN

Query Match 7.9%; Score 129.8; DB 11; Length 1112;
 Best Local Similarity 49.7%; Pred. No. 2.6e-23;
 Matches 378; Conservative 0; Mismatches 362; Indels 21; Gaps 1;

QY 671 CAACGCGGTGGCTGATGATCAGGAGATGATCCATGACGATCCTCCACGACGAC 730
 DB 18 CATCGACGGTGGCAGAGATTAATCGAGATGATTCACACTGCGGATTAATACATGATGAT 77
 QY 731 GTTATGAGCGCTCCGAGACCGAAGCAACGACATCCGGAACGAGCATTCGGAAC 790
 DB 78 GTCATGATGATAGTGGAGAGAGAGAGAGAGAACTATTCACCAACTATATGATGATCA 137
 QY 791 AAGATGCGCATTTGGCTGGTGAATTTCTTTGGAGAGCGGCGCTGCTGTCATTCGCGAG 850
 DB 138 CGGTGGCTGATCTGCTGGTGAATTTTATGTTTGCACAACTTTCTGTTCTTGGAAAC 197
 QY 851 TTGCGCAATCCGAGGTGATGAGCTTTGGCTACTGTTATTCGAAACTTGGTTGAGGA 910
 DB 198 CTAGAAATATTGAAGTTTAAATGATCAGTCAGTCAATCAAGACTTTGCAACGCGC 257
 QY 911 GAGTTCATGAGTGAATAATCTGATGATGAGGCTTGAAGGCTTAAGGAGCGAGGAA 970
 DB 258 GAGATGAACAGCTTCACCTCTTTGATGATGACATC----- 296
 QY 971 ACCTGATTAATTAATTTGAGAGACTTCTGAGAGCTGCTGCTGATTCGCAAGTCG 1030
 DB 297 ACCTGATGAGACTTCTGAGAGACTTCTGAGAGCTGATCAAGCTGATTCGAGGAGC 356
 QY 1031 TCGAGAGCAAGTGGCTTCTGAGGTGCTACCGCTGAGAGTCTGATGCTGATAGCT 1090
 DB 357 ACAAAATCAGCTTCCATATTCAGTGGCTGAGACCACTATTTGAAAAAATGATAGCA 416
 QY 1091 TACGAAAGAACTTGGTTGGATTCAGATGCTGAGACGATGCTGCACTACACGTC 1150
 DB 417 TATGGAGAGAACTTGGTCTATCTTTCAGGTTGCTGATGACATCTGATTTTCACCCAG 476
 QY 1151 TCGGTAACGACCTCGGTAAAGCCGCGGTGACAGACTTCGAGCTCGGTCTCGGCAACCG 1210
 DB 477 TACGCGGAACAATTGGCAAAACGAGCAAGTGAATTCGCAAAAGGAAACCTGATGCT 536
 QY 1211 CCGGCGCTTCTTGCATGAGAACCAACCGCGAGCTCGGTCCATGATCAAGCGCAAGTTC 1270
 DB 537 CCAAGTATCTTCTTGGCTTGAAGTGAACGAGCTTAAGGAGATCAATGATTCGAGTTC 596
 QY 1271 TCTGAGCCAGAGACGTCGAGCGTGAAGTGGTTCGAGAAAGTGAATGATTCGAG 1330
 DB 597 AGTGAACAGGATTCGTTAGCTACTGATGAACTCGTTCAATGAAGTGGTGGATTCAG 656
 QY 1331 AAGACGAGAGCTTGGCGGAGAGTATGCCAGAGGCGTTGATGCAATTCGAGAGTTC 1390
 DB 657 AAGGCAATGAGCTTGAAGAGAGAGAGGAGTTCGCAATTCGAGAGTTCGAGTTC 716
 QY 1391 CCGGAGAGTCCGCGACGAGAGGCTTTGAGACGATTCAGCA 1431
 DB 717 CCAAGAGTGAATTCAGAGTGCCTTGAAGAGATGATGATG 757

RESULT 14
 BE610448
 LOCUS 535 bp mRNA linear EST 14-DEC-2000
 DEFINITION NXSI_058_F04_F NXSI (Nsf Xylem Side wood inclined) Pinus taeda cDNA
 clone NXSI_058_F04_5', mRNA sequence.
 BE610448
 ACCESSION BE610448
 VERSION BE610448.1 GI:11779112
 KEYWORDS EST.
 SOURCE loblolly pine.
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 1 (bases 1 to 535)
 REFERENCES
 Segneroff, R. Molecular Basis of Wood Formation in the Pine Megagenome
 (Unpublished (2000))
 CONTACT Contact: Johnson, Arthur
 North Carolina State University
 Tel: 919 515 7801
 Fax: 919 515 7801
 Email: ajohnson@unity.ncsu.edu
 Seq primer: 73.
 FEATURES
 source
 Location/Qualifiers
 1..535
 /organism="Pinus taeda"
 /strain="Coastal plain loblolly pine from North Carolina"
 /db_xref="taxon:3352"
 /clone="NXSI_058_F04"
 /clone_1ib="NXSI (Nsf Xylem Side wood inclined)"
 /tissue_type="Xylem"
 /cell_type="Side"
 /dev_stage="Juvenile"
 /lab_host="XLI-Biue"
 /note="Vector: Bluescript SK; Site 1: Eco RI; Site 2: XhoI
 ; The library is from early (spring) wood, taken from
 three six-year old trees (three different genotypes), in
 the juvenile phase. These trees were induced to form side
 wood by bending to a 45 degree angle and tying them to the
 ground. Differentiating xylem was harvested from the sides
 of the inclined stems, and a mixture of all three
 genotypes was used for the library. oligo-dT primed cDNA
 was directionally cloned into the EcoRI-XhoI Bluescript SK
 vector arms. NOTE: The sequences contain a 'cDNA adapter'
 between the EcoRI site and the start of the EST. The
 adapter sequence is 'AATTCGACGAG'."

BASE COUNT 148 a 96 c 131 g 160 t
 ORIGIN
 Query Match 7.8%; Score 129; DB 12; Length 535;
 Best Local Similarity 54.3%; Pred. No. 2.6e-23;
 Matches 298; Conservative 0; Mismatches 230; Indels 21; Gaps 1;
 QY 744 CCGAGACGAGAGAAACCAACCATCCGGAACAGGCAATTCGAAACAAAGATGGCATTT 803
 DB 7 CAGATACCCGTCGAGAGTGTGGCTGTTGAACCTTGTAATGGAATAAGCTTGCTGATAC 66
 QY 804 TGGCTGATGATTTCTTGTGGACGCGGCTGTGTCATTCGAGAGTTCGCAATTCGG 863
 DB 67 TGGCTGATGATTTCTTGTGGACGCGGCTGTGTCATTCGAGAGTTCGCAATTCGG 126
 QY 864 AGGTGATGAGCTTTGGCTACTGTTAATTCGAAACTTGGTGAAGGAGATTCATGAGT 923
 DB 127 AGGTGATGAAATTAATTCGAGAGGTTGGAACATCTTGTAATGAGTGAATATGCAAA 186
 QY 924 TGAATAATTAATCTGATGATGATGATGAGCTACGCGACGAGAGAACTTCGATTA 983
 DB 187 TGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 225
 QY 984 ATTTCAGAGAGCTTACTTGAAGAGTGGCTGCTGATTCGCAAGTTCGAGAGCAAGT 1043
 DB 226 ATATCAGAGAGCTTTTACAAACAGCATCATTCATGATGATGATGATGATGATGATG 285
 QY 1044 CGCTTCTGGGTGCTACGCGCTGAGGTTGCTGATGCTTATGCTTATGCTTACGAAAGAAC 1103


```

Db 286 CTCTTATGCGGTGACCGAGAGAGTTTGCATGCTTGATATGACTATGGAAGAATT 345
QY 1104 TTGGTTTGGCATTCAGATCGTCGACGACATGCTCGACTACACCGTCTCCGCTACCGACC 1163
Db 346 TGGGATTAGCATATCAGTAGTTGATGATGCTTACTTGAATTTACTGGCACAACAGCTTCAC 405
QY 1164 TCGGTAAGCCCGGTGACAGCTCCAGTCCGCTCGCCGACCGCGCGCCCTCTTCG 1223
Db 406 TTGGCAAGGGCCCTCTTCTTGACATACGCCAGGGGATTTGTTACTGTCCAATATTGTTG 465
QY 1224 CATGGAAGACACCGCGAGCTCGGTCCCATGATCAAGCGCAAGTTCTCTGACCCAGGAG 1283
Db 466 CTTTGAAGAGTTTCCACAGTTGCATGATGATATCAGCCGAAAGTTCAAAAAGCCTGGAG 525
QY 1284 ACCTCGAGC 1292
Db 526 ATATTGATC 534

```

```

RESULT 15
BG858032      659 bp      mRNA      linear      EST 29-MAY-2001
LOCUS      1024055D07.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION      Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION      BG858032
VERSION      BG858032.1 GI:14239216
KEYWORDS      EST.
SOURCE      Chlamydomonas reinhardtii.
ORGANISM      Chlamydomonas reinhardtii
REFERENCE      Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
AUTHORS      McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.
TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model,
      Unicellular System for Analyzing Gene Function and Regulation in
      Vascular Plants; project phase 2
JOURNAL      Unpublished (2000)
COMMENT      Contact: Charles Hauser
      DCMB Box 91000
      Duke University
      Durham, NC 27708-1000
      Tel: 919 613 8159
      Fax: 919 613 8177
      Email: chauser@duke.edu.
FEATURES
      source
      1..659
      /organism="Chlamydomonas reinhardtii"
      /strain="CC-1690 wild type mt+ 21gr"
      /db_xref="taxon:3055"
      /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
      II"
      /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
      XhoI; This library, constructed by John Davies and Jeffrey
      McDermott, combines cDNAs from CC-1690 cells grown to
      mid-log phase in TAP (acetate-containing) medium in the
      light, TAP medium in the dark, HS (minimal) medium in
      ambient levels of CO2 and HS medium bubbled with 5% CO2.
      PolyA mRNA was purified from each sample, pooled and cDNA
      synthesized. The cDNA was directionally cloned into lambda
      Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
      pBluescript II SK- plasmids were excised from the lambda
      Zap clones by superinfection with ExAssist (Stratagene)
      phage. The library was normalized using method 4 described
      in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      133 a      213 c      201 g      112 t
ORIGIN

```

```

Query Match      7.6%; Score 125.4; DB 12; Length 659;
Best Local Similarity 53.7%; Pred. No. 2.7e-22;
Matches 298; Conservative 0; Mismatches 236; Indels 21; Gaps 1;

```

```

QY 670 GCNACGGCGGTTGGCTGAGTACGAGAGATGATCCATGACGATCATCTCTCCACGACGA 729
Db 125 GCACCGCGGCACTTGGCGAGATTACCGAGATGATTCACCGCGCAGCCTTGTGACACGACGA 184
QY 730 CGTTATCGACGCTTCCGAGACGAGAAACGCCACCATCCGGAACACGAGCATTCGGAAA 789
Db 185 CGTGCTCGAGAGTGCACATCCGGCGAGGACGAGACCATCAACAGCATGTATGGCAC 244
QY 790 CAAAGTGGCGATTTTGGCTGGTGATTTCTTGTGGAGACGGGCTCTGTGTGCATTTGGCGAG 849
Db 245 CGCGTGGCGGCTGCTGGCGGGGACTTCTGTTCGCCCAAGTCTCTCGTGGTTCCTCGCCAA 304
QY 850 GTTGGCGAATCCGAGGTGATTTGAGCTTTTGGCTACTGTTATTGCAAACTTGGTTGAGGG 909
Db 305 CTTGGCAACCTGGAGGTGATCAAGCTGATCAGTCAAGTGAATTCGGGACTTTGCCAACGG 364
QY 910 AGAGTTTCATGCAAGTTGAAAAAATACTGTTGATGATGCGATTGAGGCTACGGCGACGACGA 969
Db 365 CGAGATCAGCCAGCGCGCCAGCCTGTTTGACACCGACAT----- 403
QY 970 AAGTTTCGATTACTATTTCGAGAAGACTTACTTTGAAGACTGGCTCTTGAATGCCAAGTC 1029
Db 404 CACTCTGGAGCAGTACCTGGACAAGTCGTTCTACAAGACCGCCTCGCTCATCGCGCCAG 463
QY 1030 GTGCAGCAAGTGGCGCTTCTGGTGGTGGTGTGCTAGCCCTGAGGTTGCTGATGCTGCTTATGC 1089
Db 464 CTGCGCTCCGCGCGCTGTTTTCAGCGACAGCCCTGTGGAGGTGAAGGAGGCCATGTACGC 523
QY 1090 TTACGGAAGGAACCTTGGTTTGGCATTCAGATCGTCGACGACATGCTCGACTACACCGT 1149
Db 524 CTACGCAAGCACCTGGGCTGGCATTCAGGTGGTGGAGACATCTCTGGACTTACGCA 583
QY 1150 CTCGCTACCGACCTCGGTAAGCCCGCCGCTGCAGACCTCCAGCTCGGTCTCGCCACCGC 1209
Db 584 AACACGCGAGCAGCTGGCAAGCCGAGGCGCAGGACCTTGCACGCGGTAACTCAGCGC 643
QY 1210 GCGCGCCCTCTTCGC 1224
Db 644 GCCCGTCATCTTCGC 658

```

Search completed: January 16, 2003, 08:51:38
 Job time : 1707 secs

THIS PAGE BLANK (USPTO)